Escola de Camins
Escola Tècnica Superior d'Enginyeria de Camins, Canals i Ports
UPC BARCELONATECH

On the use of logistic functions for Coastal Flood Assessment

Treball realitzat per:
David Teruel Cano

Dirigit per:
Vicente Gracia Garcia
Manuel García León

Grau en:
Enginyeria Civil

Barcelona, 16 de Juny de 2017

Departament d'Enginyeria Civil i Ambiental

TREBALL FINAL DE GRAU

Abstract

Coastal flooding results as the combination of wave conditions, mean water level and terrain characteristics. The associated damage can be expressed as a function of the flooding water column and it will rarely remain stationary in a sea level scenario. The objective of the study is to determine the importance of the dynamic (waves) and hydrostatic (sea level) component as a function of a damage level and to determine what combinations of hydrodynamic drivers may lead to unacceptable damage levels.

The analysis has been performed through the use of logistic functions in the beach of Sant Pere Pescador (Girona province), covering an extension of 6,3 kilometres with an average width of 90m. The approach of the research consists of the creation of two multinomial regression models to be analyzed for 6 hydrodynamic scenarios derived from a coupled hydromorphodynamic model, one made in a node-by-node basis and the other one through polygons. The results show that a node-by-node model has nearly 10% of information loss due to the small variability and the small number of measurements. However, both models successfully predict similar expected prediction and probability to reach a flood categories for the proposed scenarios.

Results from the multinomial regression clearly reveals sea level rise as the main contribution to flood increase, with the wave return period being less significant. The model also suggests that a drastic increase in the flooding water column, approximately between 0,6 and 1 m will occur for sea level rise bigger than 0,25m. According the IPCC predictions for the sea level rise in the Mediterranean Coast, this is expected to occur between a time range of 2040-2070 (for the RCP 8.5, the most dramatic climate scenario) or starting in 2045 (for RCP 2.6 and RCP 4.5, less severe scenarios).

For what concerns the spatial extension of the flooding, the expected categories of flooding shows that the northern region of the Sant Pere Pescador beach is likely to suffer more severe flooding than the central and southern regions, both in intensity and spatial extent, reaching water depths larger than 0,6 - 1m for scenarios with sea level rise larger than 0,25m.

Key words: coastal flooding; logistic regression; sea level rise; climate change impacts

Resumen

Las inundaciones de costas son fenómenos causados por la combinación del oleaje, el nivel medio del mar y las características del terreno. El daño asociado puede ser expresado en función de la columna de agua de inundación, tomando diferentes valores para distintas situaciones de subida del nivel del mar. El objetivo del presente estudio es determinar la importancia de la componente dinámica (olas) e hidrostática (nivel del mar) como componentes de predicción del nivel de daño para poder determinar qué combinaciones de parámetros hidrodinámicos lleva a niveles de daño inaceptables.

El análisis se ha llevado a cabo a través del uso de modelos logísticos en la zona de playa de Sant Pere Pescador (Provincia de Girona), cubriendo una extensión de 6,3 Km de largo con un ancho medio de 90m. Para su desarrollo, se han elaborados dos modelos multinomiales de regresión analizando 6 escenarios hidrodinámicos. Uno de los modelos se ha centrado en el estudio nodo a nodo y el otro mediante el análisis de polígonos. Los resultados muestran que, para el modelo nodo a nodo, hay una pérdida de cerca del 10% de la información inicial debido a la escasa variabilidad en la muestra y debido al reducido número de datos disponibles. No obstante, ambos modelos consiguen predecir de manera similar la probabilidad y la predicción de llegar a ciertas categorías de inundación para los distintos escenarios.

Los resultados del modelo de regresión logística muestran la subida del nivel del mar como principal contribuyente al incremento de la cota de inundación, mientras que el período de retorno de la ola es menos significativo. El modelo también indica que para niveles de incremento del nivel del mar mayores que 0,25m, gran parte de la costa se va a ver seriamente afectada con inundaciones entre 0,6 y 1 metro de agua. Según las predicciones del IPCC, para incrementos del nivel del mar en la costa del Mediterráneo, se espera que una subida del nivel del mar alcance los 0,25m entre el 2040 y el 2070 (para el RCP 8.5, el escenario climático más severo) o 2045 (para escenarios climáticos menos graves como el RCP 2.6 y RCP 4.5).

En cuanto a la extensión de la inundación, se prevé que en la zona norte de la Playa de Sant Pere Pescador se obtengan mayores categorías de inundación, en comparación con la zona centro y sur, debido a la elevación y la configuración de las dunas. Concretamente, se esperan valores de inundación entre 0,6m y 1m en la zona norte para escenarios de subida del nivel del mar mayor que 0,25m.

Palabras clave: inundación costas; regresión logística; subida nivel mar; impactos cambio climático

Resum

Les inundacions de costes son fenòmens causats per la combinació de l'onatge, el nivell mig del mar i les característiques del terreny. Els danys associats poden ser expressats en funció de la columna d'aigua d'inundació, prenent valors variables en funció dels diferents escenaris d'augment del nivell. del mar. L'objectiu d'aquest treball és determinar la importància de la component dinàmica (onades) i hidrostàtica (nivell del mar) en la predicció del nivell de dany per tal de determinar quines combinacions dels paràmetres hidrodinàmics duen a nivells de dany inacceptables.

L'anàlisi s'ha dut a terme a través de l'ús de models logístics en la zona de platja de Sant Pere Pescador (Província de Girona), cobrint una extensió de 6,3 Km de llarg i un ample mitjà de 90m. S'han analitzat 6 escenaris hidrodinàmics mitjançant l'elaboració de dos models multinomials, un centrat en l'estudi node a node i l'altre mitjançant polígons. Els resultats mostren que, per el model node a node, hi ha una pèrdua aproximada d'un 10% de la informació inicial, degut a l'escassa variabilitat en la mostra i al reduït nombre de dades mesurades. No obstant, ambdós models aconsegueixen predir de manera molt similar la probabilitat i les prediccions d'arribar a certes categories d'inundació per a diferents escenaris.

Els resultats del model de regressió logístic mostra la pujada del nivell del mar com a principal contribuïdor a l'increment del nivell d'inundació, mentre que el període de retorn resulta menys significatiu. El model també indica que, per a nivells d'increment del mar majors de 0,25m, gran part de la costa es veurà greument afectada amb inundacions d'entre 0,6 i 1m d'aigua. Segons les prediccions del IPCC, s'espera que en la costa Mediterrània el nivell del mar pugi fins els 0,25m entre el 2040 i el 2070 (per el RCP 8.5, que és l'escenari climàtic més advers) o al 2045 (per a prediccions climàtiques menys greus, com el RCP 2.6 i 4.5)

Pel que fa l'extensió de la inundació, es preveu que a la zona nord de la Platja de Sant Pere Pescador s'hi obtinguin majors cotes d'inundació que en les zones centre i sud, degut a l'elevació i la configuració de les dunes. Concretament, s'esperen valors d'inundació entre 0,6 i 1m per a escenaris on la pujada del nivell del mar sobrepassi els 0,25m.

Paraules clau: inundació costanera; regressió logística; pujada del nivell mar; impactes del canvi climàtic

Acknowledgements

Throughout these past months, there are a group of people I would like to express my gratitude to for contributing, in one way or another, to the development of this research.

In the first place, I would like to thank my advisors Vicenç Gracia and Manel Garcia for their support during the initial stages of the project and for considering me to be part of it. Specially, I would like to thank Manel for his continuous support during the development of this project and for being always willing to discuss and debate different routes and approaches to address this project. Without our meetings, I would have not been able to delve into the world of coastal management and its complexities.

In the last but not least position, I would like to thank my family and my friends, for being always there not only during the elaboration of this research but during these past years as a Civil Engineering student.

Contents

List of Figures 9

List of Tables 12

1. Introduction 13

2. Aim and Objectives 15

3. Logistic Regression Model 16
   3.1 Introduction 16
   3.2 Linear and Logistic regression model 16
   3.3 Model Fitting 18
      3.3.1 Testing the significance of the coefficients 19
   3.4 Multiple Logistic Regression Model 21
      3.4.1 Introduction 21
      3.4.2 Multiple Logistic Regression 21
      3.4.3 Fitting the multiple logistic regression model 22
      3.4.4 Significance tests of the model 23

4. State of the Art 24
   4.1 Physical impacts of climate change on the Coastal zone 24
      4.1.1 Increase in shoreline erosion rates 24
      4.1.2 Increase of the probability and depth of flooding 25
   4.2 Coastal flooding 25
      4.2.1 Storm and Storm surges 25
   4.3 Sea Level Rise 26
   4.4 Predicting coastal flooding models 28

5. Study Area 29
   5.1 Section 20 29
      5.1.1 Characteristics of the Coastline 29
   5.2 Sant Pere Pescador Beach 30
   5.3 Study limits and zonal distribution 31

6. Methodology 32
   6.1 Coastal Flood Data set description 32
      6.1.1 Data Collection and processing 32
   6.2 Damage categorization 35
   6.3 Model fitting technique 36

7. Model results 37
   7.1 Hydrodynamic Scenarios 37
   7.2 Representation of predicted categories 39
      7.2.1 Northern region 39

7.2.2 Central region 41
7.2.3 Southern region 42
7.2.4 Comparison between Regions 42
7.3 Increase rate of categories per prediction 43
7.4 Study of the expected prediction probabilities 44
7.4.1 Complete or quasi-complete separation in Logistic Regression 44
7.4.2 Probability of Categories 46
7.5 Sensitivity Analysis 48
7.5.1 Changes in Sea Level Rise 48
7.5.2 Changes in Return Period 51
7.6 Overall discussion 52
7.7 Tipping Points and recommendations 52

8. Model evaluation and validation 55
8.1 Polygon distribution 55
8.1.1 Objectives: 56
8.1.2 Model building 56
8.2 Results of the Model 57
8.3 Validation of the Node-based model 57
8.4 Study of the expected probabilities 62
8.4.1 Comparison of node-based and polygon-based regression model 64

9. Spatial Analysis and management application 67
9.1 Study area flooding 67
9.2 Objectives 69
9.3 3D Analysis of the flooded area – Dunes as a natural barrier 69
9.4 Methodology 72
9.4.1 Data collection 72
9.4.1 Procedure 73
9.5 Determining the Assets at risk 73
9.5.1 Land Use maps 73
9.5.2 Transportation facilities 74
9.5.3 Vegetation and Land cover 76
9.5.4 Water resources affected 77
9.5.5 Building infrastructure affected 78
9.6 Risk Analysis (Application case for the Scenario 6) 79
9.6.1 Methodology of study 80
9.6.2 Risk Assessment 81
9.6.3 Determining the Consequences 81
9.6.4 Determining the Probability of Occurrence 85

10. Potential actions for Improvement 87
10.1 Introduction 87
10.2 Reduction of Risk 87
10.3 Discussion of measures 88
10.3.1 Beach nourishment 89
10.3.2 Artificial Dunes and dune rehabilitation 89

10.3.3 Set-back lines ... 90
10.4 Planning and Application ... 90

11. Conclusions ... 91
11.1 Performance of the logistic regression model ... 91
11.1.1 Validation of the Model – Node-based vs Polygon Based ... 92
11.2 Discussion on Hydrodynamic models. Sensitivity analysis Thresholds and Tipping Points ... 92
11.2.1 Concerning the Sensitivity analysis: ... 93
11.3 Assets at risk and management application ... 93
11.4 Potential recommendations ... 94

References ... 95

Annex 1: R Code

Annex 2: Maps

List of Figures

Figure 1: Nature of the condition mean of a Linear and Logistic Regression Model 17
Figure 2. Relationship of log odds and the probabilities of an event using a logistic function. ...17
Figure 3 Plot of the chi-square distribution for k = {1, 2, 3, 4, 6, 9}. Source: Wikipedia...........23
Figure 4 Simplified model of landward coastal retreat under SLR (based on the Bruun Rule) ..24
Figure 5 SLR extreme water levels and increases the probability of flooding 25
Figure 6 Schematic of a breaking wave, and wave runup, on a sandy beach with slope β. SWL is the still water line (water level without waves) and is determined primarily by the astronomical tide. R is the vertical elevation of the runup above SWL. Source: Coastal Data Information Program (CDIP) 26
Figure 7 Projections from process-based models of global mean sea level (GMSL) rise relative to 1986–2005 (Church et al., 2013) 27
Figure 8 Sea level rise projections regionalised for the period 2081-2100 with respect to the IPCC scenarios RCP 4.5 and RCP 8.5 (Losada et al,. 2014). 27
Figure 9 Conceptual diagram of flood damage functions. Source: (Yang et al., 2014) 28
Figure 10 Coastal sections defined for the Girona coastline of the coastline 29
Figure 11 Map of the study areas showing the location of the Sant Pere Pescador Beach 30
Figure 12 Coastline evolution of Sant Pere Pescador Beach from 1995, 2002 and 2004..............31
Figure 13 Zonal distribution of the study area 31
Figure 14 Representation in Zone 2 of the raster water depth map (left) and the flood area vector map (right) 32
Figure 15 Flowchart depicting the methodology adopted in this study.......................................33
Figure 16 Point Layer Detail obtained from "Point Sampling Tool" 34
Figure 17 Flood area comparison between a 2015 map (Tr:10) and a 2100 Map with a Tr:100 years. 34
Figure 18 Frequency of the categories obtained for the 5 Predictions 38
Figure 19 Selected zones for analysis 39
Figure 20 Category distribution for the 5 predictions in Zone 1 40
Figure 21 Category distribution for the 5 predictions in Zone 1, 2 and 3 41
Figure 22 Category distribution for Predictions (1), (3) and (5) in Zone 3................................41
Figure 23 Category distribution for Predictions (1), (3) and (5) in Zone 4................................41
Figure 24 Category distribution for Predictions (1), (3) and (5) in Zone 6 and 7.......................42
Figure 25 Category distribution for Predictions (1), (3) and (5) in Zone 7................................42
Figure 26 Percentage of category increase for the predictions. (a) Prediction 1; (b) Prediction 2; (c) Prediction 3; (d) Prediction 4; (e) Prediction 5 43
Figure 27 Summary of the multinomial regression for POINT 4 as shown in R 45

Figure 28 Average probability to get Categories 1-4 for the 5 Predictions....47
Figure 29 Number of points with category 1, 2, 3 and 4 under a low wave return period of 10 years for changes in Sea Level rise [0,05-0,50 m]....48
Figure 30 Number of points with category 1, 2, 3 and 4 under a low wave return period of 10 years for changes in Sea Level rise [0,05-0,50 m]....49
Figure 31 Number of points with category 1, 2, 3 and 4 under a low Sea Level Rise of 0,10m for changes in the Wave Return Period [10-75]....51
Figure 32 Number of points with category 1, 2, 3 and 4 under a low Sea Level Rise of 0,50m for changes in the Wave Return Period [10-75]....51
Figure 33 Projected regional rise in sea level up to 2100 at the nearest node of the NW Mediterranean Coast with the threshold values of Sea Level Rise chosen. Source: Own elaboration adapted from Sánchez-arcilla et al., 2016....53
Figure 34 Polygon distribution and detail of points inside each polygon....56
Figure 35 Categories obtained for Zone 3 and 4 for Prediction 1 (left) and Prediction 5 (right) 57
Figure 36 Example of Polygon 4 with the points of study inside....58
Figure 37 Comparison of expected probabilities for Category 1, 2 and 3 conditions for Prediction 1 (top) and Prediction 5 (bottom) in the North Region....63
Figure 38 Comparison of expected probabilities for Category 1, 2 and 3 conditions for Prediction 1 (top) and Prediction 5 (bottom) in the Centre Region....64
Figure 39 Comparison of expected probabilities for Category 1, 2 and 3 conditions for Prediction 1 (top) and Prediction 5 (bottom) in the South Region....64
Figure 40 Flooded Area for the 6 scenarios proposed....68
Figure 41 Transformation from the DEM model (left) to a Hillshade (right) applying the Terrain Analysis tool of QGIS....70
Figure 42 3D Representation of the Digital Elevation Model. The whole study area is represented in the top picture, and closer zoom is made to the norther region (bottom left) and the southern region (bottom right). The surface floor is given an exaggeration parameter to facilitate the interpretation....71
Figure 43 3D representation of the flooding extension under a Sea Level Rise of 0,6 m and a wave return period of 100 years. The layer of water is coloured in light blue....72
Figure 44 Flowchart showing the general methodology with the main components (hazard, exposure, flood risk and vulnerability). Source: Own elaboration adapted from (Muis, S, 2015) 73
Figure 45 Land uses map delimited for the study area with the 4 types of land uses that will be analysed: Buildings (private and public), Water Resources, Transportation roads and Vegetation and Land cover (wetlands, beach and vegetation)....74
Figure 46 Transportation routes that would be affected under Scenario 4 and 6....74
Figure 47 Km of transport infrastructure that would be affected for the 6 scenarios....75
Figure 48 Area of Land and Vegetation cover that would be affected for the 6 scenarios....76
Figure 49 Km and Area of Land & Vegetation cover that would be affected for the 6 scenarios76
Figure 50 Extension of water bodies that would be affected for the scenario 4 and 6....77
Figure 51 Km and Area of Water Bodies that would be affected for the 6 scenarios....78
Figure 52 Building Infrastructure representation with the flood area map for a Scenario 6....79
Figure 53 Flowchart of the Risk Management procedures as established by ISO 31000....80

Figure 54 Different levels of risk and the adaptation measure procedure carried out to reduce risk. (NSW Treasury, 2012) 81
Figure 55 Consequence level map resulting from the intersection of the assets at risk. 84
Figure 56 Progression of flooded area that highlights that the beach and the wetlands are the most likely flooded areas due to coastal flooding. Source: Own elaboration 85
Figure 57 Hot spot areas with a higher Risk of Flooding. Source: Own elaboration 86
Figure 58 Coastal Zone Archetype Source: RISES-AM 88

List of Tables

Table 1 Radiative Forcing of the Representative Concentration Pathways and contribution to Global Sea Level Rise in 2100. Adapted from Van Vuuren et al, 2011 and Stocker et al, 2013..26
Table 2 Total number of points per layer and the result number of extractions for the first simulation. The reference map used for intersection of the other point maps is defined with in blue. 35
Table 3 Range of flood depth values associated with a category 35
Table 4 Example point extracted from the dataframe not valid for study 36
Table 5 Predicted scenarios for the logistic model....37
Table 6 Extracted table of the category probabilities and categories obtained for Prediction 1 (SLR:0; Tr:10) 44
Table 7 Predicted scenarios for the logistic model....45
Table 8 Probability to obtain category 1, 2, 4 and 4 for the 5 predictions 46
Table 9 Main category for the predictions under changes in sea level rise and return period.....50
Table 10 Comparison between polygon and point based model for Prediction 1 60
Table 11 Comparison between polygon and point based model for Prediction 5 61
Table 12 Extracted table of the category probabilities and categories obtained for Prediction 1 (SLR:0; Tr:10) 62
Table 13 Comparison of the node-based and polygon-based model....66
Table 14 Total number of points per layer and the result number of extractions for the first simulation. The reference map defined with blue is the reference taken for the creation of the model. 67
Table 15 Total Flooded area for various return periods predicted assuming current conditions (2015) and future scenario of 0,6 m sea level rise (2100) 69
Table 16 Total Flooded area for various return periods predicted assuming current conditions (2015) and future scenario of 0,6 m sea level rise (2100) 72
Table 17 Proposed Consequences table to analyse the risk. Source: Own elaboration 83

1. Introduction

There are currently over 1.2 billion people living within 100 km of the coast and less than 100m above sea level (Purvis, et al., 2008). Coastal areas are characterised for being dynamic environments. They are one of the most valuable ecosystems in Earth, which makes them highly fragile and vulnerable to anthropogenic and natural changes.

One of the main processes affecting coastal environments is climate change (Losada, et al., 2014). Sea level rise, changes in ocean temperatures or ocean acidity are just three examples of key drivers of climate change that are expected to affect especially coastal systems in the near future. Low-lying areas, such as beaches or dunes near the sea will increasingly experience adverse consequences such as submergence, coastal flooding, and coastal erosion, mostly due to relative sea level rise (Wong et al., 2013, Boettle, et al., 2013)

The erosion of the beach and the flooding of low-lying areas are considered to be two of the processes that will be exacerbated as a combination of sea level rise and wave conditions.

Erosion during episodic storm events can be common and, in most cases, it can result with the sediments moving from the emerged region of the beach to deeper zones. However, it is only considered to be a problem when it interferes with the main functions of the beach, consisting of: a) to support system for the natural ecosystem, b) to protect the infrastructure from the waves and c) to develop a tourist activity (recreational purposes). Long term erosion in beaches occur due to increase in sea level rise and affect tons of kilometres of coast (Sánchez-Arcilla, et al., 2016).

Coastal flooding results as the combination of wave conditions (impact of the wave run-up), mean water level (affected by sea level rise) and terrain characteristics. The wave run-up is the maximum vertical extent of wave uprush on a beach or structure above the still water level (SWL) (Sorensen, 1997). The beach is considered to flood when the wave run-up has a higher height than the top height of the beach.

In combination with these risks, the Catalan littoral is also under high pressure from the attraction of tourists worldwide, that come to the Mediterranean Coast looking for the weather and the beaches. This has an impact in the use of the coastal environments and the activities that are developed. Indeed, out of the different types of coastal areas, the most vulnerable are coastal urban environments and urban beaches, because they concentrate a large number of tourists and services which reduce the space for flood

protection. In addition, urban beaches tend to concentrate large economic activity near the sea, exposing more businesses to flooding. The lack of space available for flood protection measures and the maintenance costs of the near-sea infrastructure limits the intervention to mitigate the effects of sea level rise.
In addition, apart from the exposure of urban beaches, the strong variations in the predictions for future global sea level rise brings uncertainties in how the level of future risk for socioeconomic activities would be.

Therefore, it is essential to integrate into the current coastal management procedures, adaptation strategies seeking to reduce the vulnerability to climate change and to enhance resilience of coastal systems.
But in order to apply adaptation measures, it is vital to analyse, predict and discuss how future scenarios are going to affect coastal flooding.

The protection of coastal communities relies on the ability to predict the impact of storms on sea defences. Design event will usually be a combination of many variables, including wave conditions (height, period and direction), wave and wind set-ups, tides, storm surge, river flows and foreshore height (Zou et al., 2013).

Measuring and quantifying the hazard caused by flooding is still a critical issue for coastal zone management professionals. The reason lies in the difficulties to determine the true scale at which a flooding can occur and the variability in the environmental change predictions, which may increase the magnitude of frequency of the flood hazard (Doornkamp, 1998).

The project seeks to analyse a new methodology to address and predict coastal flooding by considering the different sea level rise predictions from the Intergovernmental Panel on Climate Change report (IPCC) and changes in the wave return period.
The study has been carried out in the north-western Mediterranean Sea, in the Catalan coast, most specifically in the Sant Pere Pescador Beach. Mainly composed by fine sediment particles and shallow depth dunes, it has of over 6 kilometres of length and an average of 90 meters wide.

With this study, a combination of variations in sea level rise and wave return period have been used to predict the extension and depth of the flooding, as well as a sensitivity analysis aimed at determining for which values of SLR and wave return period the Sant Pere Pescador beach would be at risk.

Beyond the limits of the project, this research also proposes a risk analysis and a methodology to address the potential risks identified from the regression model, inviting coastal engineers, local governments and businesses in the area to integrate the results obtained in their current practices to work in making the Sant Pere Pescador beach and its surroundings a safe and functional place for everyone.

2. Aim and Objectives

The major aim of this research is to analyse and determine the importance of the dynamic (waves) and hydrostatic (sea level) component affecting coastal flooding as a function of a damage level.

Firstly, the research focuses on the creation, analysis and performance of a multinomial node-by-node regression model. To illustrate and compare the performance of this approach, a polygon-based model is also generated, built by integrating the information of some points into a given area. Both models are developed to measure the efficiency in the data processing of sea level rise and wave return period measurements. The objective is also to analyse the strengths and weaknesses of both models as potential methods for predicting coastal flooding.

Secondly, there project focuses on the real application and testing of the models in the study area with the aim to evaluate different flood hazard as a consequence of future climate scenarios. The idea behind this chapter is to:

- Analyse the main changes in the flood hazard map and the values of sea level rise and wave return period for which large flood categories would be obtained.
- Determine what combination of hydrodynamic drivers may lead to unacceptable damage levels.
- Analyse the sensitivity of logistic functions for evaluating and predicting coastal flooding for changes in Sea Level Rise and Wave Return Period.

Thirdly, the project concentrates on the spatial analysis of the assets at risk for different flood scenarios generated. In particular, the objective is to be able to measure the real implications that an increase in the water level would cause not just in the Sant Pere Pescador beach, but also in the surroundings. For this purpose, the third block aims to:

- Analyse the impacts of flooding on different land uses (transportation, vegetation, urban infrastructure...).
- Provide a qualitative methodology for the risk analysis, based on the vulnerable for different climate scenarios.
- Propose a vulnerability reduction approach that helps define an adaptation roadmap of the study area by providing a list of potential measures that could be included in an Integrated Coastal Management Plan.
- Combine coastal flood management measures with tipping points to determine the location and the time of application of the measures.

In conclusion, this project aims to evaluate a new methodology for coastal flood assessment and its integral application into the Sant Pere Pescador beach.

3. Logistic Regression Model

3.1 Introduction

Regression analysis is one of the most widely used tools to study the relationship of multifactor data. They allow to explore the relationship and behaviour between a response variable and one or more explanatory variables. The aim of an analysis using regression models is to be able to interpret how an *outcome* (dependent variable) can relate to a set of independent variables, often called *covariates*. The standard process by which this relationship is studied comprises a first step where all data is gathered, followed by the fitting of a model and an evaluation of the results using statistics (Hosmer, et al., 2003).

3.2 Linear and Logistic regression model

Out of the many different regression models for data analysis, the linear and the logistic regression model are the most widely used.
The main difference between them is the nature of the outcome variable. For linear regression, it is assumed to be continuous while for the logistic regression model is binary or dichotomous.

For any regression model, one key element is the conditional mean, expressed as $E(Y|x)$. It is the mean value of the outcome variable, given a certain value of the independent variable. It can also be described as the expected value of Y, given a certain x.
In the case of a linear regression, this mean is defined as an equation linear in x:

$$E(Y|x) = \beta_0 + \beta_1 x \tag{1}$$

The expression allows the conditional mean to range from $-\infty\ to + \infty$.

For the case of a dichotomous outcome, the conditional mean must hold values between 0 and 1 (Figure 1):

$$0 \leq E(Y|x) \leq 1$$

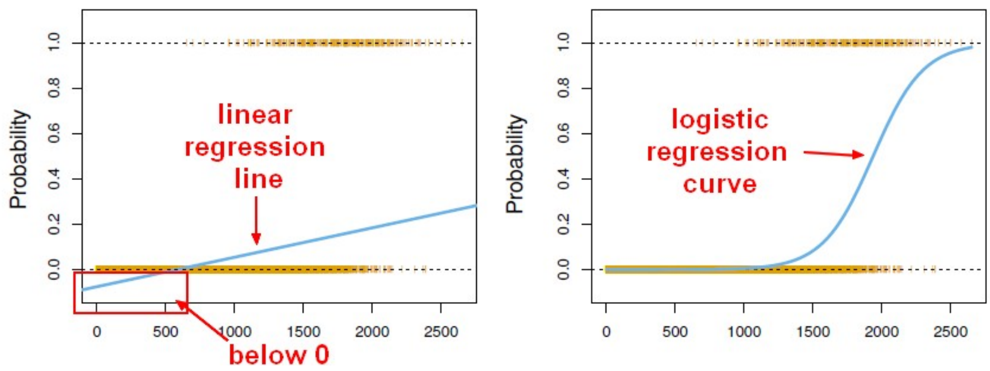


Figure 1: Nature of the conditional mean of a Linear and Logistic Regression Model

As it can be seen from Figure 1, the curve shown is said to be S-shaped, or sigmoid curve. For this type of plots, the changes of $E(Y|x)$ per unit change of x reduces progressively as the condition mean gets close to 0 or 1.
There are several distribution models apart from the logistic model that can be used to represent $E(Y|x)$ in the situation where Y is dichotomous. However, for this project, a logistic distribution will be considered because it has a flexible and good mathematic use of the function and because model parameters provide the basis for meaningful estimates of effect. Additionally, the sigmoid curve is a better alternative than the linear function when the modelled random variable has a bounded upper/lower limit. In that regard, the linear function is unbounded and it can lead to unrealistic values for extreme values of a given random variable.

Therefore, the notation of the conditional mean given the logistic distribution is determined by:

$$\pi(x) = E(Y|x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \tag{2}$$

To facilitate the study of the logistic regression, a logit transformation of the conditional mean $\pi(x)$ can be applied, resulting in:

$$g(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x \tag{3}$$

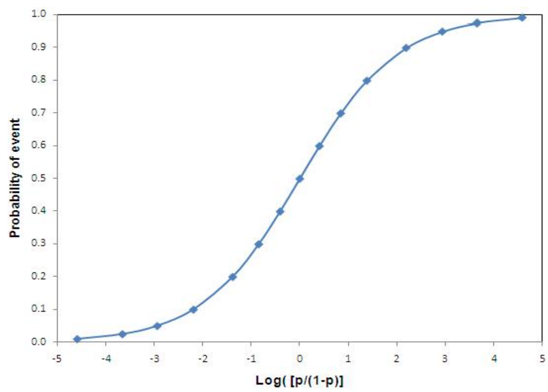


Figure 2. Relationship between log odds and the probabilities of an event of occurring using a logistic function.

The logit transformation shares several properties with the linear regression model such as linear in its parameters, continuous and can range from $-\infty\ to + \infty$.

Between a linear and a logistic regression model, there is also a difference in the conditional distribution of the outcome variable:
For a linear regression model, the observation of the outcome variable is assumed as:

$$y = E(Y|x) + \varepsilon \tag{4}$$

with ε being the error, or the deviation from the conditional mean. The error follows a normal distribution with mean zero and with a constant variance.
For a logistic regression, this is not the case and the outcome variable given x is:

$$y = \pi(x) + \varepsilon \tag{5}$$

In this situation, the error can take two possible values. When $y = 1$, the error is $\varepsilon = 1 - \pi(x)$ and when $y = 0$, the error is $\varepsilon = -\pi(x)$. Therefore, the error has a distribution with mean zero and variance equal to $\pi(x)[1 - \pi(x)]$.

Apart from the differences between linear and logistic models both in the limits of the conditional mean for a logistic model and the differences in the distribution of the error, the techniques used in the model fitting of a linear regression follow the same principles as those from the logistic regression.

3.3 Model Fitting

In order to fit the logistic regression model presented in equation (3), it requires the estimation of the unknown parameters β_0 and β_1.
For logistic regression, a method that can be used is the maximum likelihood (Wilks, 1938). This method estimates the beta coefficients by finding the values that maximize the likelihood of making the observations given the parameter.

To apply the method, it is first needed to compute the **likelihood function.** The maximum likelihood estimators will be the values obtained by maximizing this function. Supposing a sample of n independent observations of the pair (x_i, y_i), being y_i the dichotomous outcome variable and x_i the independent variable associated to the ith subject. The outcome variable is assumed to be 0 or 1 representing the presence or absence of a phenomenon. As an example of how this would be like in this project, 0 would mean that there is no flooding; whereas 1 would mean the opposite.
Considering a pair (x_i, y_i), if $y_i = 1$, the contribution to the likelihood function is $\pi(x_i)$ and for $y_i = 0$, it would be $1 - \pi(x_i)$. Therefore, for the pair (x_i, y_i), the contribution to the likelihood function can be conveniently expressed as:

$$\pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \tag{6}$$

Considering the observations as independent, the likelihood function can be obtained multiplying the different observations:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{n} \pi(x_i)^{y_i}[1 - \pi(x_i)]^{1-y_i} \tag{7}$$

As established by the maximum likelihood principle, the estimates of $\boldsymbol{\beta}$ that will be used are the ones that will maximize the function *(7)*. In order to obtain those values, the equation is transformed with a log:

$$L(\boldsymbol{\beta}) = ln[\,l(\boldsymbol{\beta})] = \sum_{i=1}^{n} \{y_i \ln\,[\pi(x_i)] + (1 - y_i)\ln\,[1 - \pi(x_i)]\} \tag{8}$$

The values of $\boldsymbol{\beta}$ are obtained differentiating the equation (8) with respect to β_0 and β_1, which results in two equations, called likelihood equations:

$$\sum_{i=1}^{n} [y_i - \pi(x_i)] = 0 \tag{9}$$

$$\sum_{i=1}^{n} x_i[y_i - \pi(x_i)] = 0 \tag{10}$$

For logistic regression, both equations are nonlinear in β_0 and β_1 and they require an iterative method to find the solution of the equations.
The $\boldsymbol{\beta}$ values obtained from solving the equations (9) and (10) are called the maximum likelihood estimates and are denoted by $\widehat{\boldsymbol{\beta}}$. Similarly, the $\hat{\pi}(x_i)$ corresponds to the maximum likelihood estimate of $\pi(x_i)$, and provides an estimate of the conditional probability that Y is equal to 1, representing the "fitted" or predicted value for the logistic regression model.

As a consequence of equation (9), another equation can be obtained:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\pi}(x_i) \tag{11}$$

which means that the sum of the observed values of y is equal to the sum of the predicted (expected) values.

3.3.1 Testing the significance of the coefficients

Once the logistic model is fitted, the significance of the variables in the model should be assessed. This is carried out through the formulation and testing of statistical hypothesis for checking whether the independent variables have a significant relation to the outcome variable. In this sense, part of this process is not only to check the goodness of the fit, but also checking whether the predicted values are an accurate representation of the

observation. The method aims to check if the covariate is significant and if the predicted values are better or not with the variable inside the model.
For this reason, the model fitting goal is to answer the following question: *Does the model that includes the variable in question tell us more about the outcome (or response) variable than a model that does not include that variable?*

The evaluation for the logistic regression model consists of the comparison between the observed values of the response variable and the predicted values obtained from models, both for the cases when the variable in question is included or not.
The comparison is made through the use of the likelihood function and a saturated model (the one that contains as many parameters as data points are present). It is based on the following expression:

$$D = -2\ln\left(\frac{likelihood\ of\ the\ fitted\ moded}{likelihood\ of\ the\ saturared\ model}\right) \tag{12}$$

with the likelihood ratio being the value inside brackets.

Since the distribution of Equation (8) is known, putting it inside the expression above, allows to be used for hypothesis testing purposes. This test is defined by the **likelihood ratio test** (Neyman and Pearson, 1933) and is represented by the deviance, a quality-of-fit statistic for a model often used for statistical hypothesis testing that is represented as:

$$D = -2\sum_{i=1}^{n}[y_i\,ln\left(\frac{\hat{\pi}_i}{y_i}\right) + (1-y_i)ln\frac{1-\hat{\pi}_i}{1-y_i}] \tag{13}$$

where $\hat{\pi}_i = \hat{\pi}(x_i)$.

As mentioned before, in order to assess the significance of the independent variable, a comparison can be made with the value of D with and without the independent variable in the equation. The change of D due to including the independent variable in the model is:

$$G = D(model\ without\ the\ variable) - D\ (model\ with\ the\ variable) \tag{14}$$

G is a statistic that behaves similarly to what the numerator of the partial F-test does in linear regression. It can also be expressed as:

$$G = -2\ln\left(\frac{likelihood\ without\ the\ variable}{likelihood\ with\ the\ variable}\right) \tag{15}$$

Developing the expression, G can also be expressed as:

$$G = 2\left\{\sum_{i=1}^{n}[y_i \ln\,[\pi(x_i)] + (1-y_i)\ln\,(1-\pi(x_i))] - [n_1 ln(n_1) + n_0 ln(n_0) - nln(n)\right\} \tag{16}$$

where:

- the first term is the log- likelihood $L(\boldsymbol{\beta})$
- $n_1 = \sum y_i$
- $n_0 = \sum(1 - y_i)$

There are also two relevant statistical tests to validate the obtained parameters: The Wald Test and the Score test (Hosmer, et al., 2003). For the present study and since the Score test is not available in many software packages, it won't be considered. The Wald Test has been chosen as the metric.

The Wald test is equal to the ratio between the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$ and the estimate of its standard error. It can be expressed as:

$$W = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \tag{17}$$

3.4 Multiple Logistic Regression Model

3.4.1 Introduction

In many case study scenarios, the outcome is not always dependent on one variable, leaving the prediction conditional to a subset of variables to be examined. The Multiple logistic regression model, as in the case of linear regression, can handle many independent variables.
The following chapter will present how the multiple logistic regression model is built, the fitting and the tests for the significance of the model.

3.4.2 Multiple Logistic Regression

We consider a collection of *p* independent variables grouped by the vector $\boldsymbol{x}' = (x_1, x_2, \dots, x_p)$. Considering the conditional probability that the outcome takes place, $Pr(Y = 1|\boldsymbol{x}) = \pi(\boldsymbol{x})$, the logit of the multiple logistic regression model can be expressed by:

$$g(\boldsymbol{x}) = \ln\left(\frac{\pi(\boldsymbol{x})}{1 - \pi(\boldsymbol{x})}\right) = \beta_0 + \beta_1 x + \beta_2 x_2 + \cdots + \beta_p x_p \tag{18}$$

where,

$$\pi(\boldsymbol{x}) = \frac{e^{g(x)}}{1 + e^{g(x)}} \tag{19}$$

This formulation assumes that the variables are scaled as intervals. For the cases where the variables are discrete, it is not appropriate to include them in the model as in the abovementioned formula. Instead, the expression of the variables considering that there a p-variables and in the j_{th} there is discrete, it would be:

$$g(\boldsymbol{x}) = \beta_0 + \beta_1 x + \sum_{l=1}^{kj-1} \beta_{jl} D_{jl} + \cdots + \beta_p x_p \tag{20}$$

In general, if there are k values of a nominal scaled variable, the required design variables would be k-1.

3.4.3 Fitting the multiple logistic regression model

Similar to what it has been done in the invariable case, in order to fit model, the estimates of the vector $\boldsymbol{\beta}' = (\beta_0,\ \beta_1, \dots, \beta_p)$ need to be calculated.
The method of estimation will be the maximum likelihood. The result equation obtained is:

$$\sum_{i=1}^{n} [y_i - \pi(\boldsymbol{x_i})] = 0 \tag{21}$$

and

$$\sum_{i=1}^{n} x_{ij}[y_i - \pi(\boldsymbol{x}_i)] = 0 \tag{22}$$

For j=1,2,...,p

In the multivariate case, the equation is very similar to what it is obtained for one variable with the difference that equation $\boldsymbol{\pi(x_i)}$ is now (19).

In order to solve the above likelihood equation, it is necessary to use software available in nearly every statistical software package. The solutions estimates are denoted by $\widehat{\boldsymbol{\beta}}$ and the fitted values for the multiple logistic regression model are $\widehat{\boldsymbol{\pi}}(\boldsymbol{x}_i)$**.**

The estimation of the standard errors of the estimated coefficients $\widehat{\boldsymbol{\beta}}$ is carried out by estimating the variances and covariances through the theory of maximum likelihood estimation, which states that the estimators are obtained from the matrix of the second partial derivatives of the log-likelihood function:

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2} = -\sum_{i=1}^{n} x_{ij}{}^2\, \pi_i(1 - \pi_i) \tag{23}$$

and

$$\frac{\partial^2 L(\beta)}{\partial \beta_j^2 \beta_l} = -\sum_{i=1}^{n} x_{ij} x_{il}\, \pi_i(1 - \pi_i) \tag{24}$$

for j,l=0,1,2,...,p

The variance and covariance of the estimated coefficients are obtained as:

$$Var(\beta) = \mathbf{I}^{-1}(\boldsymbol{\beta}) \tag{25}$$

Where $\mathbf{I}(\boldsymbol{\beta})$ is the *observed information matrix (p+1) x (p+1)* that contains the negative terms given in (*23*) and (*24*).
The estimator of the variances and covariances, denoted as $\widehat{\mathrm{Var}}(\hat{\beta})$, are obtained by evaluating $Var(\beta)$ at $\hat{\beta}$.

The estimated standard errors of the estimated coefficients, which will be used in developing methods for coefficient testing and confidence interval estimation can be denoted as:

$$\widehat{\mathrm{SE}}(\hat{\beta}_j) = [\widehat{Var}(\hat{\beta}_j)]^{1/2} \tag{26}$$

for j=0,1,2,...,p

3.4.4 Significance tests of the model

To describe and evaluate the significance of a model, multivariate logistic regression uses also the same methods as described for one variable logistic models and it is based on the statistic G. However, it presents some variations in the fitted values, $\hat{\boldsymbol{\pi}}$, that in the multivariate case are based on the fitted model that contains p+1 parameters $\hat{\boldsymbol{\beta}}$.
Under the assumption of a null hypothesis that the p coefficients for the covariates in the model are equal to zero, the distribution of G is a chi-square with p degrees of freedom.

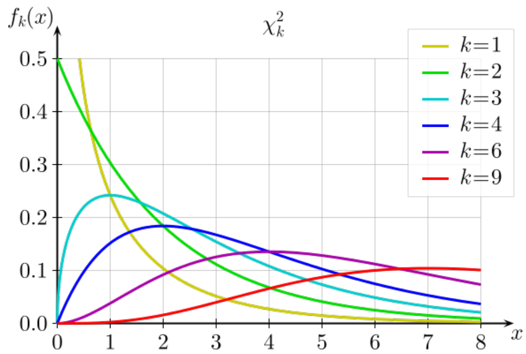


Figure 3 Plot of the chi-square distribution for values of k = {1, 2, 3, 4, 6, 9}. Source: Wikipedia

If one or more p coefficients are different from zero, the null hypothesis can be rejected. However, before concluding that any or all of the coefficients are nonzero, the Wald test statistic for univariable can be checked:

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \tag{27}$$

After the statistics have been checked and since the goal is to find the best fitting model while minimizing the number of parameters, the next step is to fit a reduced model only working with the variables thought to be significant and compare the results obtained with the first model. Depending on the differences between both models, it can be decided whether there is enough statistical justification for including or not certain variables in the model.

4. State of the Art

4.1 Physical impacts of climate change on the Coastal zone

There are many impacts associated to climate change, but from a human perspective, the five most important effects of climate change in the coastal zone are: increased probabilities of (1) coastal flooding; (2) coastal erosion; (3) rising water tables; (4) saltwater intrusion into surface and groundwater and (5) biological effects (Klein et al., 2006).

4.1.1 Increase in shoreline erosion rates

Erosion is the physical movement of sediment away from the shore via wave and current action. Due to climate change, SLR increase will have the capacity to erode larger surface of the shore by promoting offshore transport of sediment (Zhu et al., 2010).
Although very simplified, the 'Bruun Rule' offers a model of the erosion process by explaining that the shoreline recession is in the range between 50 to 200 times the rise in relative sea level and is caused so as to maintain an equilibrium beach profile.

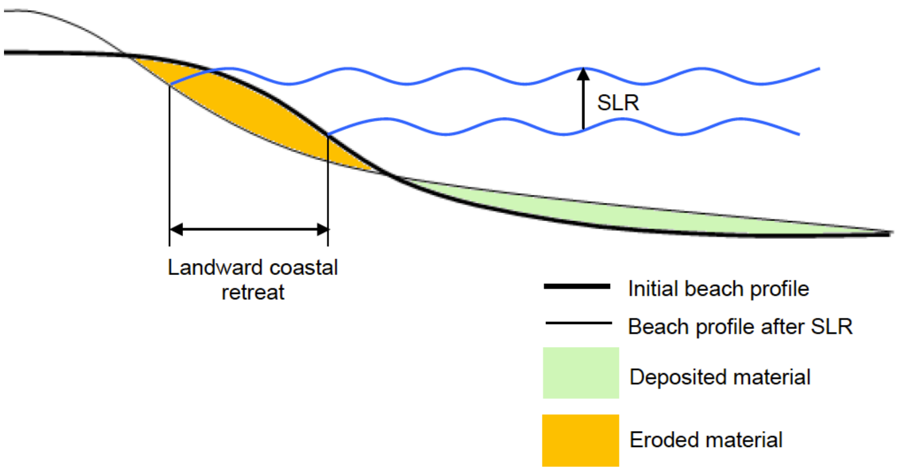


Figure 4 Simplified model of landward coastal retreat under SLR (based on the Bruun Rule) (Zhu et al., 2010).

In a climate change scenario, with an increase in the Sea Level Rise, the sediment is removed from the beach causing erosion, and subsequently deposited offshore, with the nearshore zone gaining elevation at an equal rate to the sea level rise. While the SLR increases, the profile of the beach adjusts landwards and upwards by removing sediment from the beach and depositing in the nearshore zone (and maintaining equal the volume of eroded and deposited material).

4.1.2 Increase of the probability and depth of flooding

Sea Level Rise can increase the probability of flooding in coastal areas, as illustrated in Figure 5, linking high storm water levels with the capacity to cause greater damage.

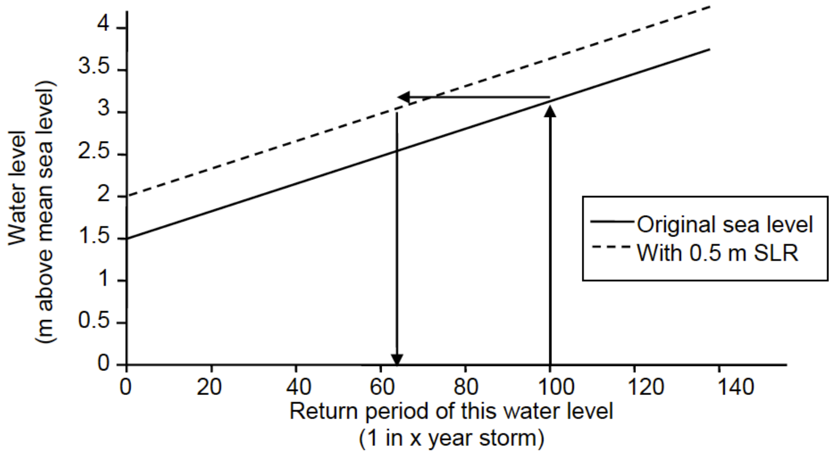


Figure 5 SLR raises extreme water levels and increases the probability of flooding without adaptation

Having more frequent extreme water levels will be aggravated by the degradation of natural coastal systems such as marshes and dunes which currently serve as natural coastal defences. For many coastal communities, natural defences are vital to protect from extreme events (Zhu et al., 2010).

4.2 Coastal flooding

Low elevation areas, such as coastal environments, suffer from fluvial and coastal flooding. Depending on its duration and intensity, the combination of both events can lead to serious increase in the flooding depth. Bearing in mind that this project aims to determine the impacts associated to climate change, only the flooding caused by coastal dynamics will be considered.

Coastal flooding results from the combination of wave conditions, mean water level and terrain characteristics. Some events have a special impact in flooding:

4.2.1 Storm and Storm surges

Storm surges can increase water level and are formed as a result of low atmospheric pressure and extreme winds. The level of flooding is also affected by high tide levels that, combined with storms, can increase the water level near the coast (NOAA, 2015).

Large waves during high tides can lead to an erosion of beaches, resulting in problems to the main functionalities of the beach. This process can be caused by the *wave run-up* (Figure 6), which is the maximum vertical extent of wave uprush on a beach or structure above the still water level (SWL) (Sorensen, 1997). As it has been seen, erosion could be caused by directly impacting the beach, removing material, or redistributing it to the foreshore and nearshore.

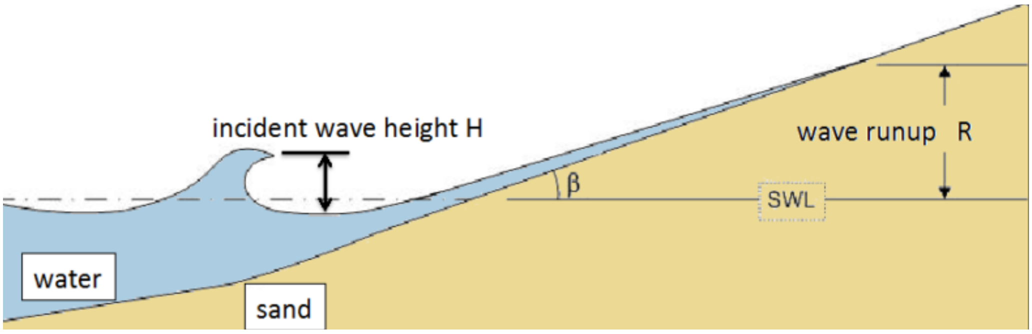


Figure 6 Schematic of a breaking wave, and wave runup, on a sandy beach with slope β. SWL is the still water line (water level without waves) and is determined primarily by the astronomical tide. R is the vertical elevation of the runup above SWL. Source: Coastal Data Information Program (CDIP)

Apart from the above-mentioned causes, an increase of coastal flooding can occur due to larger scale regional and ocean scale variations which result as a consequence of ocean dynamics and the seasonal heating and cooling of the ocean.

4.3 Sea Level Rise

One important component of this project relies on the effect that changes in Sea Level rise will cause to the coastline. Since there is uncertainty over possible sea level scenarios, the range of magnitudes that will be used in the analysis are extracted from the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) *Climate Change 2013: The Physical Science Basis.*

The IPCC Report establishes 4 different scenarios, defined by Representative Concentration Pathways (RCPs) that vary depending on the total radiative forcing (cumulative measure of human emissions of GHGs from all sources expressed in Watts per square meter) (Church et al., 2013). Table 1 presents the median values and likely ranges for projections of global mean sea level (GMSL) rise and its contributions in metres in 2100 relative to 1986–2005 for the four RCP scenarios:

Description		Global Sea Level Rise in 2100 (m)	
		Mean	Likely range
RCP8.5	Rising radiative forcing pathway leading to 8.5 W/m2 in 2100.	0.74	[0.52 to 0.98]
RCP6	Stabilization without overshoot pathway to 6 W/m2 at stabilization after 2100	0.55	[0.38 to 0.73]
RCP4.5	Stabilization without overshoot pathway to 4.5 W/m2 at stabilization after 2100	0.53	[0.36 to 0.71]
RCP2.6	Peak in radiative forcing at ~ 3 W/m2 before 2100 and decline	0.44	[0.28 to 0.61]

Table 1 Radiative Forcing of the Representative Concentration Pathways and contribution to Global Sea Level Rise in 2100. Adapted from Van Vuuren et al, 2011 and Stocker et al, 2013.

One of the stages of the project will involve an estimation of the contribution of sea level rise scenarios to coastal flooding. For this purpose, it is worth noticing that sea level rise is expected to increase in all scenarios, even in the most optimistic case of RCP2.6, which assumes a limitation in the increase of global mean temperature below 2°C (Figure 7).

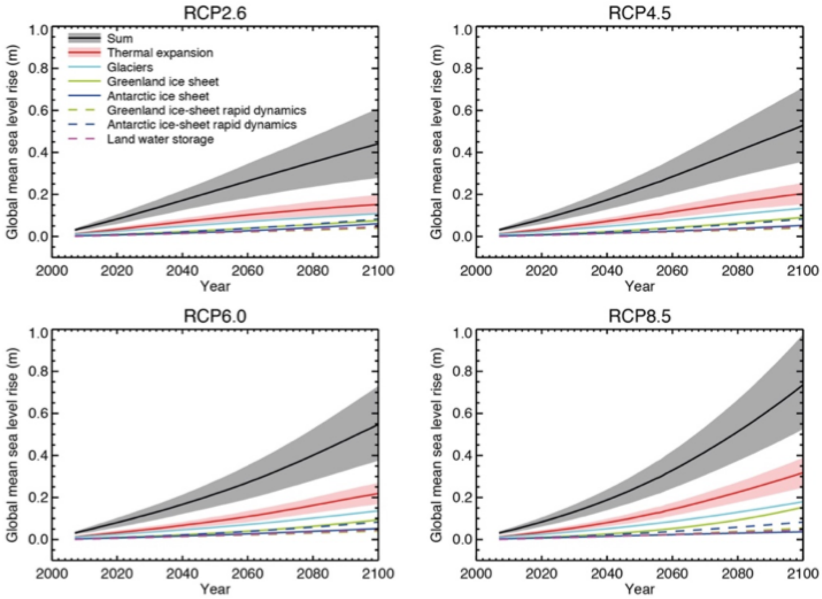


Figure 7 Projections from process-based models of global mean sea level (GMSL) rise relative to 1986–2005 (Church et al., 2013)

However, it is important to consider that the values presented are referred to global means. Past changes suggest that there are local variations of the sea level. Changes in sea level are not uniformly distributed around the world due to, for example, ocean circulation (Generalitat de Catalunya, 2016).

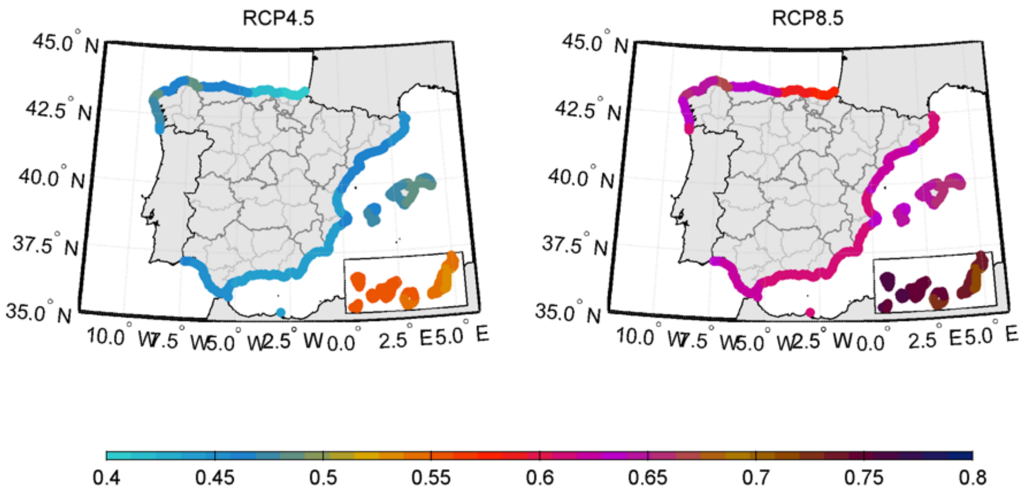


Figure 8 Sea level rise projections regionalised for the period 2081-2100 with respect to the IPCC scenarios RCP 4.5 and RCP 8.5 (Losada et al,. 2014).

The projections for the Mediterranean Sea show that the increase in the sea level will be 10% less than the global one, which means that, from the global levels, 10cm of difference should be accounted for (Figure 8).

4.4 Predicting coastal flooding models

Flood damage functions are among the most common methods for flood damage estimation worldwide (Yang et al., 2014).
Normally, for the analysis of flooding in a river basin, flood functions are defined by relating flood severity (as measured by depth, volume, duration, etc.) with projections of the resultant damage in a specific area, which are usually derived using historical flood damage information (Yang et al., 2014).

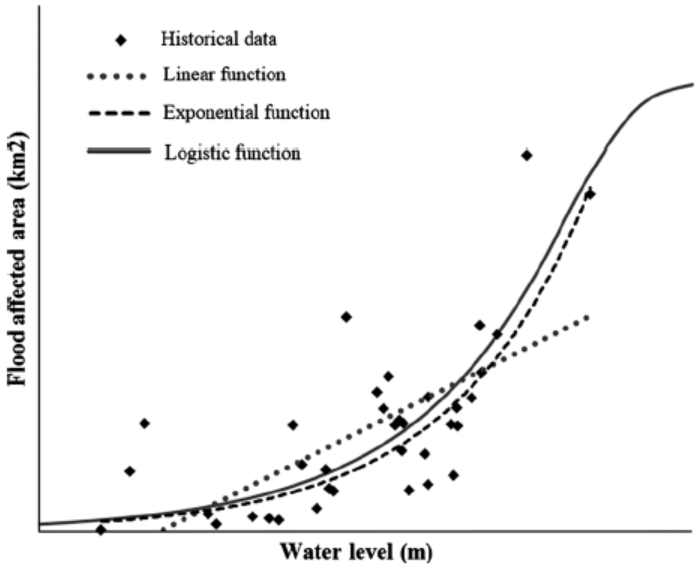


Figure 9 Conceptual diagram of flood damage functions with observed flooded area. (Yang et al., 2014)

There are many different function shapes that can vary depending on the study, being the logistic, exponential, and linear functions three of the most commonly used (Figure 9).

In coastal areas, there are several variables that can be used to predict coastal flooding. One way to proceed is to use a digital elevation model together with current or future predictions of Sea Level Rise in order to track the areas that would be flooded under certain scenarios. More complex models include the use of terrain characteristics and the impact of waves and storm surges to account for the variability and the dynamics of the coastal environment (Sánchez-Arcilla et al., 2014). In the end, it is always important to bear in mind that there are several components affecting coastal flooding and choosing one variable or another one will influence the potential duration and intensity of floods.

5. Study Area

The project area is defined along the Catalan coast, that is part of the north-western Mediterranean Sea. More specifically, the study is carried out in the Sant Pere Pescador beach, located in the province of Girona. The Green Book of the Coastal Conditions of Catalunya (CIIRC, 2010) divides the coastline of Girona in 7 sections, with the area of study located in section 20 (Figure 10).

5.1 Section 20

Nearly half of the beaches in this section are located in urban environments, 35% in natural environment shared with urbanisations and nearly 20% are occupied by camping. Half of the beaches have a seafront promenade and also infrastructure such as sports clubs or access roads. Some beaches have suffered from surface loss due to heavy storms. Infrastructure damages were detected also in the study beach of Sant Pere Pescador, where also sediment management operations have been carried out in multiple occasions.

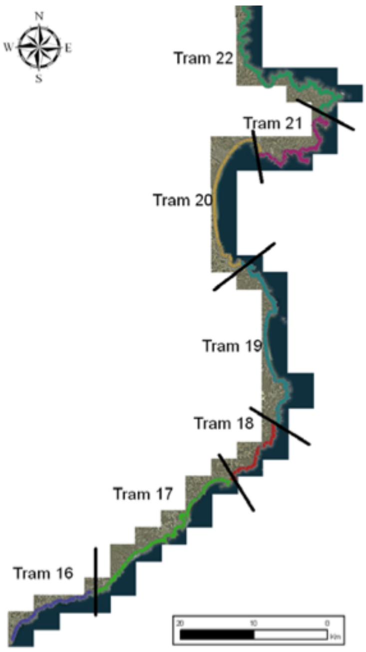


Figure 10 Coastal sections defined for the Girona coastline of the coastline (CIIRC, 2010)

5.1.1 Characteristics of the Coastline

There are several aspects that define the area of study than can be helpful to explain the behaviour and expose of the beach to flooding:

Sediment size:
The northern extreme of sector 19 and sector 20 register the smallest size of sediments of the Girona coast, with a medium value of d_{50}=0,34 mm.

Coastline evolution
Most of the Girona coastline is classified as an erosive coast, with 80,05% of its beach extension retreating. The average erosion along the whole Girona coastline is -1,5 meters/year. Due to the morphology of the study area, characterised by long sandy beaches, the Sant Pere Pescador and its adjacent beaches register the biggest rates of erosion, reaching an average value of -2,2m/y.

As a consequence of this, sectors 19 and 20 register 83% of the total sediment loss of the province of Girona (CIIRC, 2010). This is also because large sediment movement is associated with fine sediment areas such as section 20. The coast of the province of Girona is characterised by having beaches between cliffs (i.e. pocket beaches) and they area more stable than Sant Pere Pescador beach.

Figure **11** Study are showing the location of the Sant Pere Pescador Beach. Source: Google Earth

5.2 Sant Pere Pescador Beach

The Sant Pere Pescador beach (Figure 11) is mainly formed by fine sediment particles and shallow depth dunes. The length of the beach is 6,3 kilometres, is ˜90 metres wide and is located 1,5 km distance to the urban area. The extension of the area is strongly limited with rock cliffs, which have a strong influence in the characteristics of the region.

Sant Pere Pescador Beach	
General Information	
Length (km)	6,3 km
Average Width (m)	88,35 m
Area (m^2)	564459,43
Average Slope	0,10
Sediment diameter (mm)	0,23 mm
Beach orientation (º right with respect north)	175
Hydrodynamics	
Average wave Hs (m)	0,61
Frequent direction	0 N 27%
Sea level range (m)	0,44
Sea level with T_R 100 year	0,75
Environmental Issues	
Type of environment	Urban / Natural / Beach
Area included in INUNCAT	No
River presence	Yes

Dunes	Yes
Humid areas registered	Yes
Uses	
Main use type	Tourism / Recreational
Other uses	Sports
Port	No
Occupancy level	Medium
Infrastructures	Roads / Sport School
Other spaces	Showers, bars, sport material renting, etc.

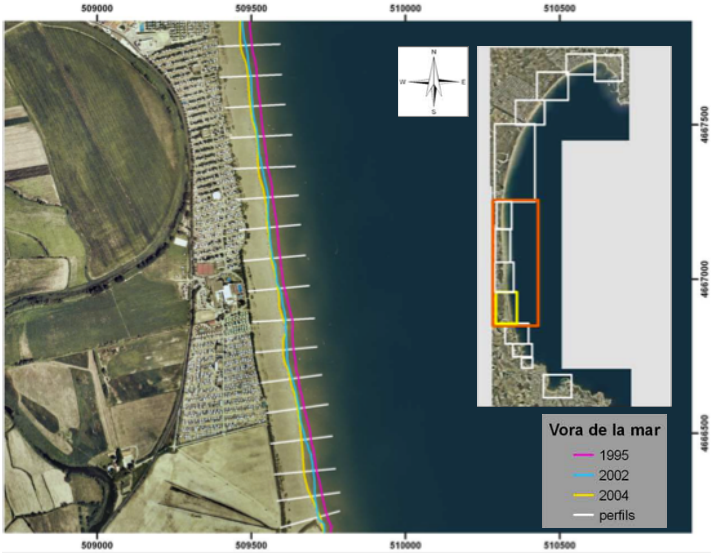


Figure 12 *Coastline evolution of Sant Pere Pescador Beach from 1995, 2002 and 2004* (CIIRC, 2010)

5.3 Study limits and zonal distribution

Since the study area is very large in the Y-axis, it has been decided to divide the whole beach in 8 zones (Figure 13).

This distribution will ease the analysis and comparisons that will be made during the project. Each zone has been named in accordance to its position, with Zone 1 located in the north of the Sant Pere Pescador beach, and Zone 8 located in the south. Zone 1 limits in the north with Empuriabrava and the beginning of the Can Comes Beach. The end of Zone 1, Zone 2 and half of Zone 3 includes the remaining part of the Can Comes Beach.
The centre part of the study area, which is described from part of Zone 3 until Zone 6 is where Sant Pere Pescador Beach is located.

The remaining Zone 7 and 8 include the Devesa Beach and limits in the south with Sant Martí d'Empuries.

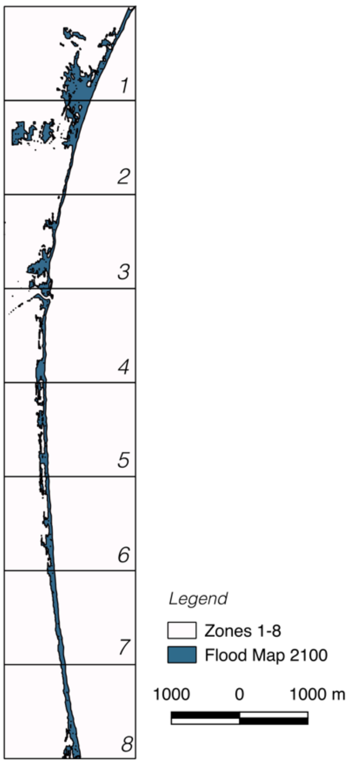


Figure 13 Zonal distribution of the study area

6. Methodology

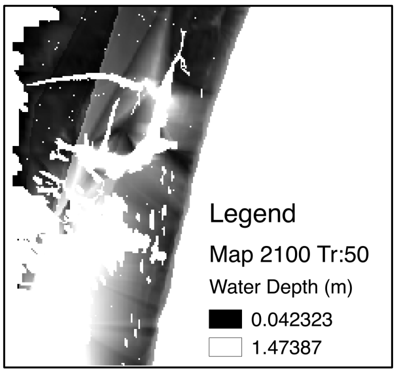
6.1 Coastal Flood Data set description

In order to assess the efficiency and robustness of a multiple logistic regression model to predict coastal flooding, first it is required to adapt the variables needed from the input data. Figure 15 describes the process that has been followed to obtain the data frame.

6.1.1 Data Collection and processing

The input data consists of:

- **Water Depth Raster Layers**: 6 raster maps that show the flooded depth values for the study area in a 2015 and 2100 scenario. For each year, three different maps have been obtained for scenarios with 10, 50 and 100 year-return period.
- **Vector Flood Area maps:** 6 polygon maps indicating the flooded area, taken from the contour of the raster maps are also given. (Figure 14).



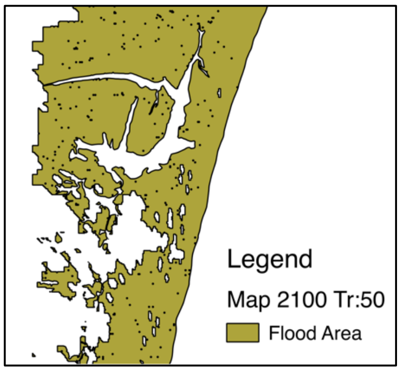


Figure 14 Representation in Zone 2 of the raster water depth map (left) and the flood area vector map (right)

The given maps are obtained from a study carried out between the *Institut Cartogràfic I Geològic de Catalunya* (ICGC) and the *Laboratori d'Enginyeria Marítim*a of the Polytechnic University of Catalonia (LIM-UPC) (Pinyol et al., 2015). The project developed a numerical model combining hydrodynamic and morphodynamic models in order to determine potential flooding areas under a set of extreme scenarios.

The project analysed the impact of three types of erosion for sand beaches: the episodic, the medium-term and the long term. For this project specifically, long-term erosion results have been used.

In order to obtain the maps of Figure 14, Pinyol et al., 2015 carried out the following procedure:

1. Definition of the wave conditions for 10, 50 and 100 wave return periods. The scenarios are defined for an increase in the Sea Level of 0,6 m (RCP 8,5) as expected in year 2100;
2. For each return period, different wave directions were associated depending on the orientation of the beach. At the same time, each surge direction had one storm intensity depending on the registered values of the closest buoy;
3. The wave propagation was then modelled from open waters until a 15m water depth through the SWAN model (Booij et al. 1999)
4. The flooding area caused by the wave was modelled through a morphodynamic model XBEACH (Roelvink et al. 2009).

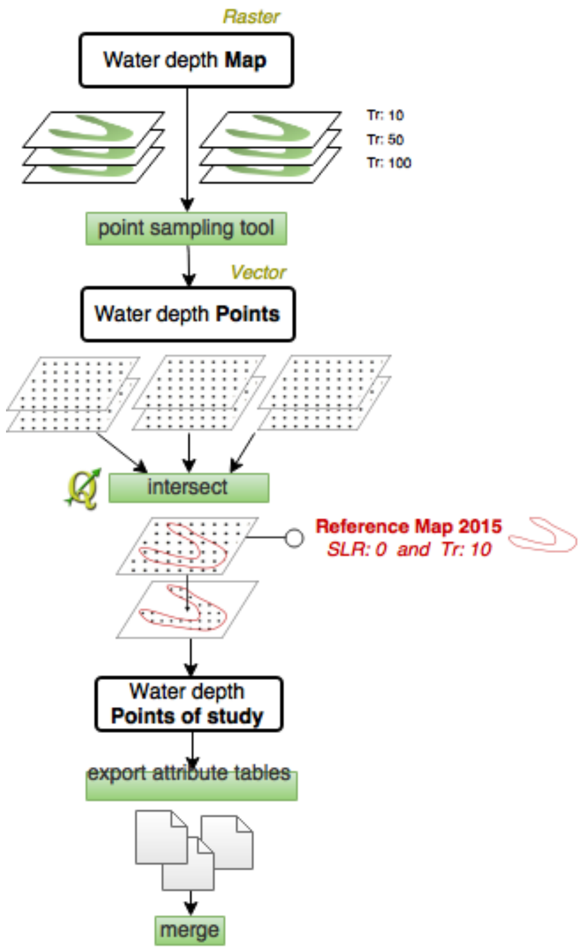


Figure 15 Flowchart depicting the methodology adopted in this study

It is important to highlight some characteristics of the variables of these input maps. For the baseline scenario at 2015, it is assumed a Sea Level Rise (SLR) of 0m, since it will be the reference point that will be used for the comparison with future scenarios. Also, the maps that show the predicted water depth in 2100 are calculated considering a Sea Level rise of 0,6 meters, which is inside the possible range of values that the IPCC suggests. As it was mentioned in Chapter 4, the value taken from the RCP 8,5 refers to the Mediterranean prediction.

The first step is to set the layer's coordinate reference system (CRS) of the Raster Flood Maps in QGIS using the Universal Transverse Mercator (UTM) 31N ETRS89. The map is then converted into vectorised point layers, which means that each pixel of the raster file will be associated to a vector node containing the water depth for a specific return period and sea level rise. To obtain the point layers, the "*Point sampling tool*" in QGIS is used.

After the process is done, the output files obtained are vector maps represented by points (Figure 16).

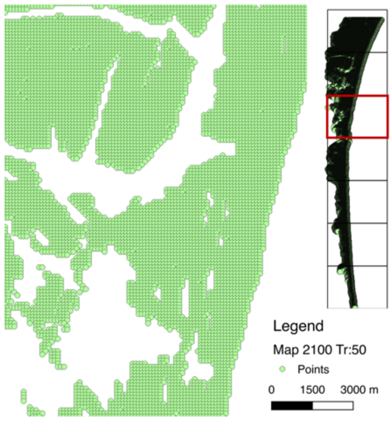


Figure 16 *Point Layer Detail obtained from "Point Sampling Tool"*

As it can be seen in Figure 17, there is a big difference between the flooded area of 2015 with a return period of 10 years and the map of 2100 with a return period of 100.
The first stage of this study will adjust a logistic regression in each of the points obtained from the raster map, which means that the aim is to find points with as much information as possible. Taking a node from the 2100 map (with a return period of 100), for example, since it is the most critical scenario, the information on a specific point far from the coast will only be representative in that scenario, because for less catastrophic scenarios, the point would not be flooded. This would limit the amount of information available for the construction of the logistic regressions. By choosing a 100 year-return period, it is likely that most of the maps do not flood the same extension of area, resulting in points with no data.

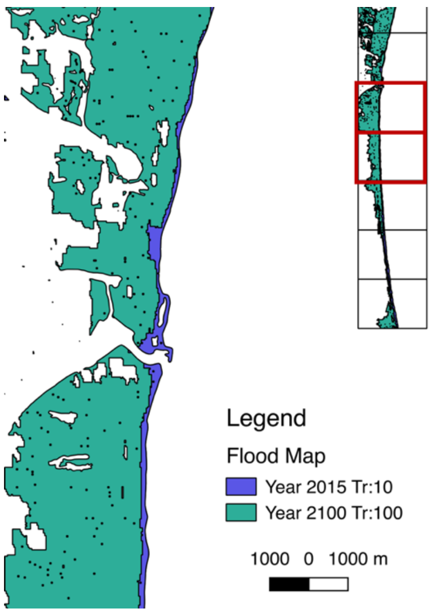


Figure 17 Comparison of the flood area for a 2015 map with Tr:10 and a Map of 2100 with a return period of 100 years.

It is for this reason that first, all the points that are common and suffer from flooding in the 6 scenarios have to be selected. For this purpose, it is required to intersect all the point maps with the reference map.

The reference map is a 2015 scenario with 10 year-return period. It is the map with less extension of flooding, so by choosing it as a reference point, all the points from the other maps will have flooding data available. By intersecting all the maps with the reference one allows to find that all the points that will be analysed are be the ones under the same perimeter.

In order to obtain only the data of the points inside the reference map, the "Intersect" geoprocessing tool in QGIS has been used. The total number of points of each layer and the result of points extracted is detailed in Table 2. As it can be seen, this first analysis ignores a great amount of points that are located outside the reference map. The accuracy and performance of the logistic model with this restriction in the number of measurements will be an issue that will be discussed during the whole study.

	Conditions		Total Number of points	Number of points extracted
	SLR	Return Period		
2015	0	10	3.928	3.928
		50	5.764	
		100	9.212	
2100	0,6	10	15.593	
		50	54.920	
		100	64.000	

Table 2 Total number of points per layer and the result number of extractions for the first simulation. The reference map used for intersection of the other point maps is defined with in blue.

6.2 Damage categorization

Once the information of the observations and the 2100 projections maps have been obtained, and before performing a statistical analysis of this information, the flooding depth has to be associated to a certain damage category.

The global mean sea level rise is expected to range from 0,09 to 0,88 in 2100 according to the Third Assessment Report of the IPCC. Not only the SLR but many other parameters involved in coastal flooding are expected to change, with high uncertainty in their values. Since the objective of this project is not to predict a precise value of flooding depth that depends on different scenarios, but rather suggest which is the probability to reach a certain level of damage, and to measure if changing the associated parameters related to coastal flooding scenarios would have a big or small impact in the overall damage, the values of flood depth taken from the map have been converted to 4 categories as described in Table 3:

Category	Flood depth (m)
1	0 – 0,4 m
2	0,4 – 0,6
3	0,6 – 1
4	> 1

Table 3 Range of flood depth values associated with a category

It is worth to mention that the chosen categories are not fixed. Depending on the type of project and the nature of the variables, the number of categories can vary, being the most common choices from 3 to 5. For this project, having 3 categories would overestimate the effects of flooding, associating small values of water depth to high categories. Using 5 categories would widen too much the spectrum of possible water depth. Since the possible range of water depth values expected goes from 0 to more than 1 and due to the project constraint, that there aren't many measurements available, having 5 categories would split too much the data, making it more difficult to determine the general trend and the effectiveness of using logistic regression in coastal flooding.

6.3 Model fitting technique

First-level analysis

From the initial map layers, a dataframe of 3928 points/per map has been obtained, each point with 6 scenarios. As it was previously discussed, all the points included in the dataframe have a certain value of flood depth, even in the map where there is less probability of flooding. This is because the base (or reference) map used to extract the points has the minimum amount of points flooded, which assures that all the other maps will contain a certain value for the flooded point.

Data processing and the statistical analysis of the multinomial logistic regression is conducted using the programming language R, with the software package (*nnet*).

There is an important consideration to be taken before fitting the multinomial regression. Since there is a small amount of measurements, some points might have only one measurement of a category (Table 4).

POINTID	GRID_CODE	POINT_X	POINT_Y	HS_TR	SLR	Category
8	0,60334897	510177,075	4675712,366	10	0	3
	0,98513597	510177,075	4675712,366	10	0,6	3
	0,714647	510177,075	4675712,366	50	0	3
	0,91110402	510177,075	4675712,366	50	0,6	3
	0,818039	510177,075	4675712,366	100	0	3
	0,95719099	510177,075	4675712,366	100	0,6	3

Table 4 Example point extracted from the dataframe not valid for study

In this example, point number 8 has only 1 category 3 which makes it impossible for the model to fit a logistic regression and therefore, to converge. It is for this reason that, before calculating, the data frame needs to be filtered. Therefore, the points that don't satisfy the requirement to have more than one measurement of a certain category have been excluded.

From the total 3928 points of study, there were 296 rejected and 3632 points valid for study. Given the fact that in this first part of the study, only 6 scenarios are taken per point, having a reduction in the number of points of study by 7,5% is an important disadvantage for the precision of the model, since some data available won't be used.

With the final dataframe values, the model can be computed, predicting the category of hazard as a consequence of sea level rise (SLR) and the return-period (Tr_Hs).

$$|Category \sim SLR\ (Sea\ Level\ Rise) + TR\_HS\ (Return\ Period)|$$

7. Model results

7.1 Hydrodynamic Scenarios

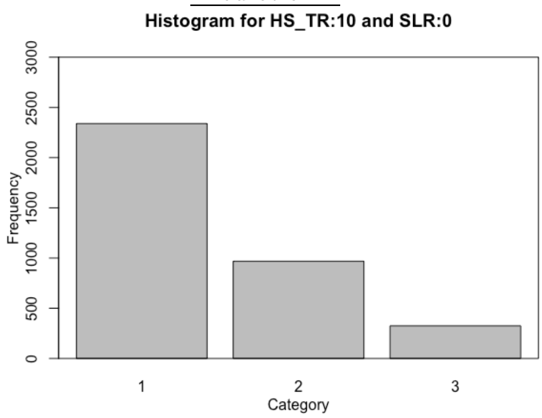
The logistic regression model is assessed through the prediction of 5 scenarios with changes in the return period and the sea level rise. Those values have to be between the range of values used when building the model, in order to be able for the model to predict accurately. The values of return period given were 10, 50 and 100 and, for the sea level rise, two scenarios were conceived: one with 0 m sea level rise based which represents current ad 0,6 m which is one medium projection for 2100.
Therefore, the 5 scenarios that have been tested with the model can be separated by:

Prediction	Sea Level Rise (m)	Return Period (year)
1	0	10
2	0	25
3	0	50
4	0,6	75
5	0,6	100

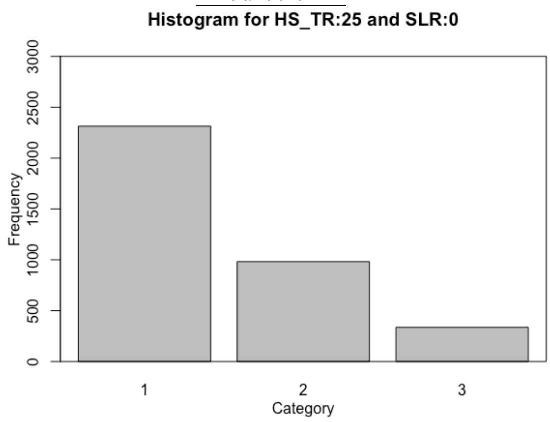
Table 5 Predicted scenarios for the logistic model

Plotting the histograms for the 5 predictions already gives an idea of the main category for each case:

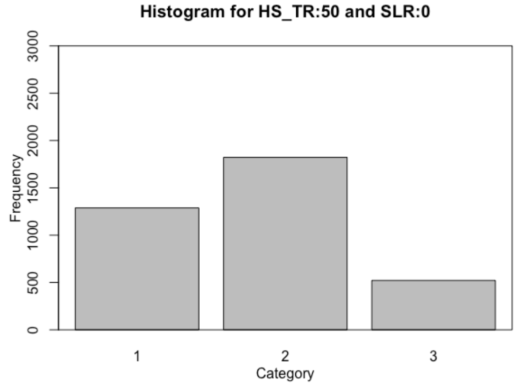
Prediction 1



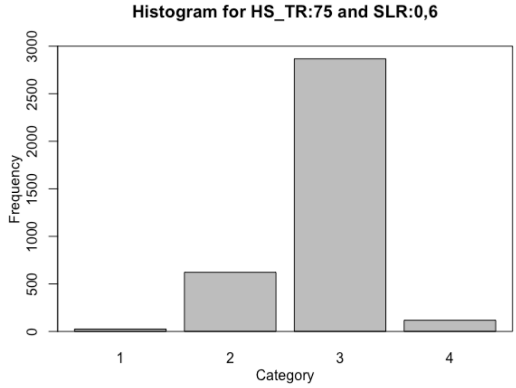
Prediction 2



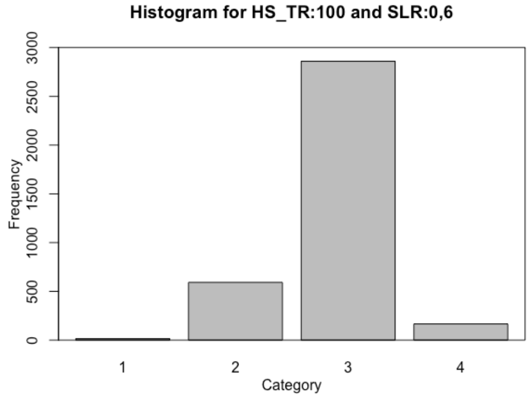
Prediction 3



Prediction 4



Prediction 5



Prediction	**Frequency**			
	1	**2**	**3**	**4**
1	2339	968	325	0
2	2314	982	336	0
3	1288	1822	522	0
4	25	622	2867	118
5	15	592	2859	166

Figure 18 Frequency of the categories obtained for the 5 Predictions

For the first scenarios with 0 m in Sea Level Rise, small changes in return period of 10 and 25 do not seem to highly contribute to an increase in the Category. For both cases, the major Category is 1, with water depths varying from 0,1 to 0,4 m. Prediction 1 counts 2339 points in Category 1 while 2314 points are included for Prediction 2. The only difference between both Predictions is that the latter one has more points in Category 2 and 3. Therefore, the distribution of categories observed Predictions 1 and 2 remain very similar. In both cases, they do not register any point in Category 4.

It is not until a 50-year return period is plotted that the most probable category surpasses Category 1. Prediction 3 has a reduction in more than 1000 points in Category 1. Those points seem to have shifted to Category 2, that registers over 1800 points, the double from Prediction 2. This might be because most of the points from Prediction 2 that were inside Category 1 had values of water depth close to the limit between Category 2 and 3. In a smaller scale, Prediction 3 registers also an increase of points inside Category 3, getting over 500 points.

Points in Category 4 are found only once a sea level rise of 0,6m is accounted, as in Prediction 4. As the histogram shows, the transition from Prediction 3, where there was a return period of 50 and a SLR of 0m and Prediction 4 is that the vast majority of

points are in Category 3, reaching x5 the number of points obtained in Category 3 from Prediction 3. Also, there are no significant points in Category 1, where there is a drastic reduction from 1288 to 25 from Prediction 3 to 4.
Lastly, there is not much difference between Prediction 4 and 5, where there is only an increase of the return period from 75 to 100. For the last prediction, there is an increase in the number of points in Category 4.

There are two issues that arise from the analysis of the histogram. First, how much more does sea level rise contribute in the prediction of the damage category compared with the return period and, secondly, how does this contribution change over the different scenarios. Does sea level rise become more determinant for small values of return period or the opposite? How relevant is the sea level rise?

Overall, it can be seen that the tendency of the histogram is to swift to higher categories with increases in sea level rise and return period. This confirms the initial assumption and common belief that with more sea level rise there is an increase in the water depth. In addition, from the histograms it can also be seen that the sea level rise seems to be the variable that contributes most. This has been observed in the transition from Prediction 3 to Prediction 4, where the change in return period was the same as in the different predictions but the sea level rise was significant (increasing from 0 to 0,6m).

7.2 Representation of predicted categories

To assess coastal flooding, not only is important to determine how much does the category increase per predictions but also how is this increase spatially distributed.
Understanding how each region in the study area is affected by the morphodynamic and hydrodynamic variables is key to provide the most useful management measures.
In order to represent the changes in category, the predicted categories extracted from *R* have been assigned to the point layers in QGIS.
Due to the large length of the study area, the analysis will be carried out in three specific zones separated between them, in order to compare how the changes in category take place in the North, Centre and South of the study area. The zones for analysis are represented in Figure 19.

Figure 19 Selected zones for analysis

7.2.1 Northern region

The northern region is represented by Zone 1 2 and 3. Figure 20 shows the density of categories and its location for different scenarios in Zone 1. As it can be seen, for Prediction 1, 2 and 3, the number of points in category 1, 2 and 3 is similar, with Prediction 3 already having a large amount of points in category 3. The big difference is seen for Prediction 4 and 5, where all the points are either category 3 or 4. As it can be seen, the most exposed areas for zone 1 would be the top right corner and the bottom left. This spatial analysis of the evolution of categories is very useful because, for future analysis, checking from satellite images and the elevation could give further explanation of why those areas suffer more the impacts of coastal flooding. It is also an alternative to provide specific management measures to mitigate the impacts of flooding in those specific points.

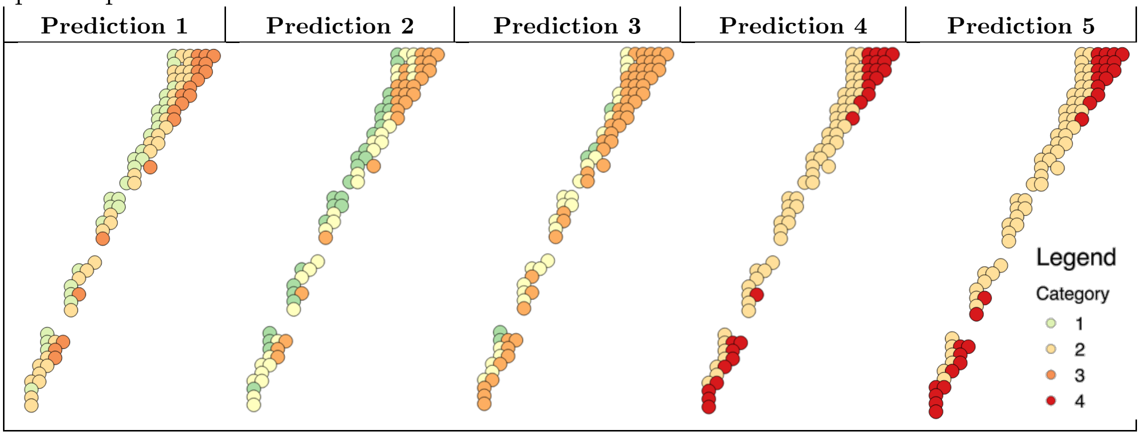
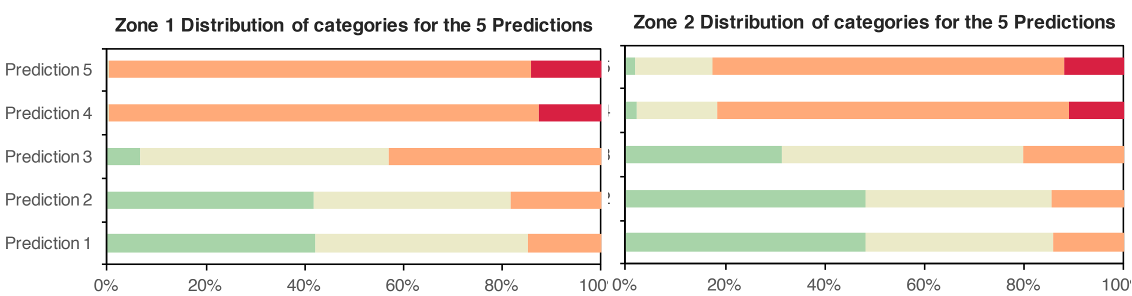


Figure 20 Category distribution for the 5 predictions in Zone 1

Alternatively, if the percentage of category per prediction is represented for the three zones (Figure 21), it can be seen that there is no big difference between the three zones. From the three zones, number 1 is the most exposed, not only because for small predictions it registers the smallest amount of category 1 and 2 points, but also because for more severe consequences, as established in Prediction 4 and 5, the zone registers nearly 1 80% of points in category 3, with the rest in category 4. Conversely, Zone 2 seems to be slightly less affected by the predictions, since it is the only zone from the three that has category 1 and category values in the highest predictions.

At this stage, it can not be yet explained what might have been the reason behind the differences between the zones. It could either be due to the topography differences of each zone analysed, which could act as a barrier or not to coastal flooding. However, those differences in the categories obtained could also happen due to the model itself, and have no direct relation with the topography. It is for this reason that a further study will be carried out in the following chapters.

What it can be said is that, in general, for Predictions 1 to 3, the vast majority of points are comprised between category 1 and 2, whereas for higher predictions the major category is 3.



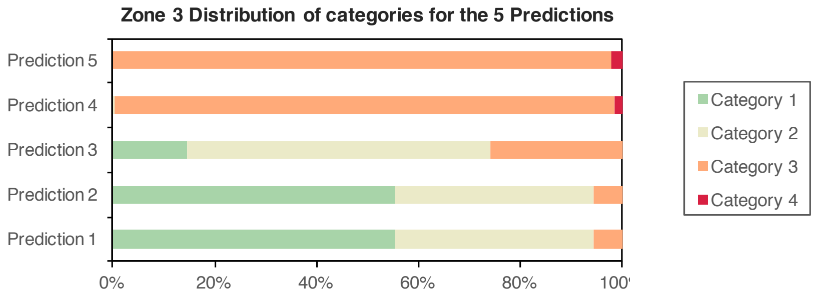


Figure 21 Category distribution for the 5 predictions in Zone 1, 2 and 3

7.2.2 Central region

The central region seems to be less susceptible to coastal flooding compared to the Northern region. This can be seen from a big amount of points having category 1 water depth for small Predictions 1 and 2. In addition, for larger Predictions like 4 and 5, even though category 3 continues to be dominant throughout the whole region, for both zones there are values of water depth of category 2 (opposed to what was seen in the Northern Region), as well as nearly no points in category 4. Between Zone 4 and Zone 5 there are no major differences, apart from the fact that for prediction 4 and 5, there seems to be a slight increase in the number of category 2 values compared to Zone 4.

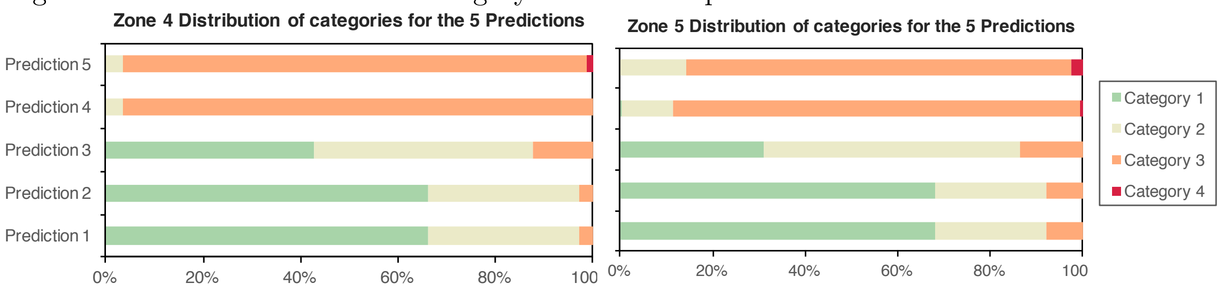


Figure 22 Category distribution for Predictions (1), (3) and (5) in Zone 3

Evaluating how does the increase in category evolve throughout the different Predictions Figure 23 also show local increases in category, close to the coast (right). The local maxima that affects certain points from flooding can be easily identified by looking at the red points from Prediction 5. As it can be seen, even for Prediction 1 and 3, those regions are also the ones that have the highest category level of the whole map. In general, the points suffering from higher flooding are the ones on the right, most exposed to the coast.

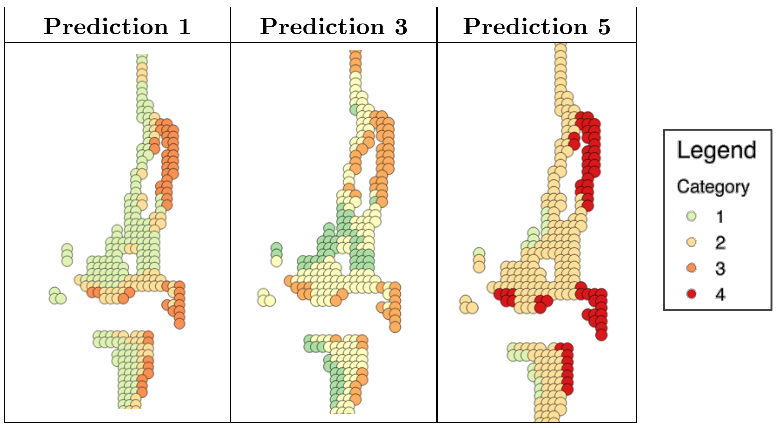


Figure 23 Category distribution for Predictions (1), (3) and (5) in Zone 4

7.2.3 Southern region

The Southern region has very similar results to the ones obtained for the central region Figure 24, with more than 70% of the points from Prediction 1 and 2 in category 1. For Prediction 3, both Zones register a similar number of points in category 1 and 2, with more than 40% of the points already in category 2.
The biggest different between this region and the central region is accounted for Predictions 4 and 5. In this zone, close to 40% of the points are in category 2 (for Zone 6) and more than 30% for Zone 7. However, those values are significantly higher, since for central region, the proportion of category 2 points was less than a 20%. This is an indication that in the southern region as well, there is less exposure to coastal flooding than in the Centre and North.

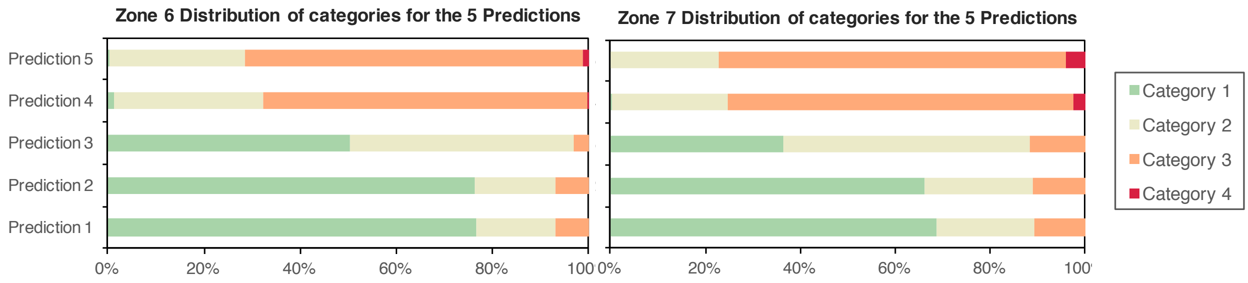


Figure 24 Category distribution for Predictions (1), (3) and (5) in Zone 6 and 7

Figure 25 represents the distribution of categories for Zone 7 and Prediction 1, 3 and 5. The figure shows that the coastline is also the most exposed for the three scenarios, with higher categories reached. In addition, there are, also for this region, local points where the increase in category and water flooding is higher than the rest (the red dots in Prediction 5).

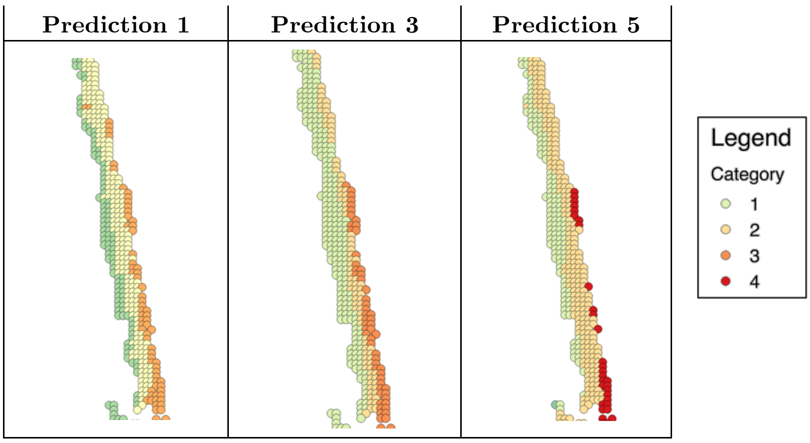


Figure 25 Category distribution for Predictions (1), (3) and (5) in Zone 7

7.2.4 Comparison between Regions

Overall, the spatial analysis and distribution of categories can be a very useful tool to complement with a quantitative analysis of the results, by giving details on which regions and zones are more exposed to coastal flooding. In general, a trend has been identified and is that, the more north it is analysed, the higher are the number points in higher categories. This means that the northern region is more vulnerable to coastal flooding and is more impacted than the centre and southern region.

This information will be very useful when analysing the potential risks associated with each zone of the study area and when discussing the potential implications that coastal flooding would cause to the coastline. As it has been seen, some areas are more vulnerable than others.

7.3 Increase rate of categories per prediction

One important aspect to consider when discussing the main category of each prediction, is also how many points increase category in a more drastic scenario. As Figure 26 shows, in the first prediction, only 1% of the points have an increase in the category. That can represent that, for small increments of return-period, the changes in water depth is not very large. Also, it can mean that return period for small values is not a significant measure of coastal flooding. However, there is a 34% of increase registered in the scenario with SLR:0 and return period of 50. This big increase is very interesting because it can help establish a threshold value for which there is a significant increase in the category level.

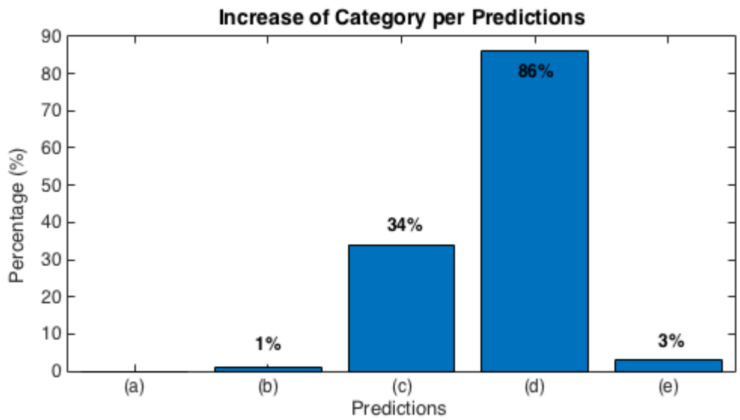


Figure 26 Percentage of category increase for the predictions. (a) Prediction 1; (b) Prediction 2; (c) Prediction 3; (d) Prediction 4; (e) Prediction 5

With still a sea level rise of 0, the fact that there has been a fairly big increase in the category means that for return periods smaller than 50, the threshold of water depth is close to the limits between categories stablished in the damage categorization.
For climate adaptation, information like this one could help establish tipping points, suggesting that for return periods close to 50, measures should be taken to prevent reaching a higher category level in the area.
Reinforcing the idea already explained before, a prediction of an increase in the sea level rise significantly increases the points that reach a higher category level, with more than 85% reaching higher levels of category. Comparing this value with the previous 34% proves that sea level rise is the main variable influencing coastal flooding.
Lastly, comparing the prediction of a sea level rise of 0,6 and a return period going from 75 to 100, it can be seen that, changes in return period don't alter the category levels, since there is only an increase in the category of 3%.

In order to fully understand the impact of sea level rise in coastal flooding, a more detailed study should be carried out with predictions for different sea levels ranging from 0 to 0,6.

7.4 Study of the expected prediction probabilities

Analysing the outcome probabilities can give more details on how to model works and how does it behave for different simulations. It can be very useful because, opposite to the category study, the probabilities show also which are the expected probabilities of the categories that are not dominant.
Table 6 shows a fragment of the resulting probabilities obtained through R for the Prediction scenario 1:

POINTID	**Category Predicted**	**Probability**			
		Category 1	Category 2	Category 3	Category 4
1	1	0,999959545	2,39175E-05	1,65377E-05	0
2	2	0	5,58511E-05	0	0
3	2	0	0,999336097	0,000663808	9,50891E-08
4	3	0	0	0	0
5	3	0	0	1,99392E-05	0

Table 6 Extracted table of the category probabilities and categories obtained for Prediction 1 (SLR:0; Tr:10)

Since this project works most of the times with small amounts of information, there are different outputs that this table presents and that depend on the information each point has had in the building of the multinomial model. For example, as it can be seen, for some points such as POINTID=2 or 5, the model only associates one probability to the category that is dominant.
For other points like POINTID=1 or 3, the model shows different probabilities associated to a certain category. Those are the most interesting points for this study because not only show how is the category going to be in a certain scenario, but also can give a hint on what is the expected probability going to be, in a near future, under a more aggressive scenario.
Taking POINTID=1 again as an example, it can be seen that, although the obtained Category is 1 shown, there is an existing probability, even though it is smaller, to reach Category 2. This suggests that, in a scenario of, for example, more Sea Level Rise, POINTID=1 would be more exposed to suffer a higher category whereas points that only show a unique probability of Category 1 are more likely to stay in that category.

Identifying where those points having more spreading probabilities are located could be very beneficial for future work by the government, since it would give an accurate description of the category of flooding in a specific area.

7.4.1 Complete or quasi-complete separation in Logistic Regression

As it can also be seen from Table 6, there are some points such as POINTID=4 that don't show any probability, which given the fact that it predicts a certain category, the results obtained is clearly nonsense. This is due to a quasi-complete or complete separation.

Separation occurs if the predictor (or a linear combination of some subset of the predictors) is associated with only one outcome value when the predictor is greater than some constant (IDRE, 2017)

The occurrence of complete separation in practice depends on the sample size, the number of subjects with the outcome present, and the number of variables included in the model. The sensitivity of the overlap fitness depends on the sample size and the range of the covariate. The very large estimated coefficients and especially the large estimated standard errors can give a hint on which variable is the dominant (Hosmer et al., 2003).

The numerical problems of a zero cell count due to complete separation are manifested by spurious large estimated standard errors and sometimes by a large estimated coefficient as well (Hosmer et al., 2003).

```
> summary(fit)

Call:
multinom(formula = Category ~ HS_TR + SLR, data = x.sub, maxit = 3000)

Coefficients:
                  Values  Std. Err.
(Intercept) -30.5791517 356.932851
HS_TR         0.4130478   5.137396
SLR          62.0976951 634.584184

Residual Deviance: 0.0001820405
AIC: 6.000182
```

Figure 27 Summary of the multinomial regression for POINT 4 as shown in R

Taking into account the summary of the multinomial regression of POINT 4 (Figure 27) It can be seen that the estimated standard errors of the Sea Level Rise are much larger than it would be expected. This suggests that the estimates might not be stable for some points and that separation could be a problem.
For the scope of this project, the interest relies on how the model represents and reproduces future coastal flooding. This is why, for the analysis of the probabilities, the points that suffer from complete or quasi-complete separation will be extracted.

Prediction	Total points of study	Number of complete separation points	Final points of study
1		435 (12%)	3197
2		451 (12,5%)	3181
3	3632	435 (12%)	3197
4		20 (0,5%)	3612
5		20 (0,5%)	3612

Table 7 Predicted scenarios for the logistic model

By extracting the points of study that have complete or quasi-complete separation, it can be seen that predictions 1 to 3 are the ones with higher percentage of. This means that predictions with small values of sea level rise and return period, the model is less useful to predict the category and the probability of reaching a certain category because one variable (as it was seen, Sea Level Rise), can solely predict the outcome (without the contribution of the return period).
For scenarios of high sea level rise (SLR=0,6) and high return periods, it can be seen that there are less points where complete separation occurs. In particular, and as the table shows, for predictions where the sea level rise is 0,6m (Prediction 4 and 5), the percentage of points that deal with complete or quasi-complete separation drops from 12% to 0,5%. This suggests that in order to obtain the categories for higher prediction scenarios, it has been necessary to use the logistic regression model whereas for small Prediction scenarios, and possibly due to the fact that the predictor is associated with only one outcome value, more points have suffered complete or quasi-complete separation. In these high Prediction scenarios, even considering that the sea level rise has a stronger influence in the multinomial regression, the variable return period has a higher weight in the total prediction than in the logistic regression of Prediction 1, 2 and 3.

7.4.2 Probability of Categories

Once the points that don't follow the multinomial model are extracted, the probabilities obtained from R can be analysed. Table 8 gives the measured probabilities to obtain each category for the different 5 predictions:

Prediction	Category (%)			
	1	2	3	4
1	88%	6%	7%	0%
2	78%	15%	7%	0%
3	90%	7%	3%	0%
4	0%	18%	78%	4%
5	0%	17%	78%	5%

Table 8 Probability to obtain categories for the 5 predictions

The results have been represented in Figure 28 to study the tendency and patterns that the different predictions might have between them:

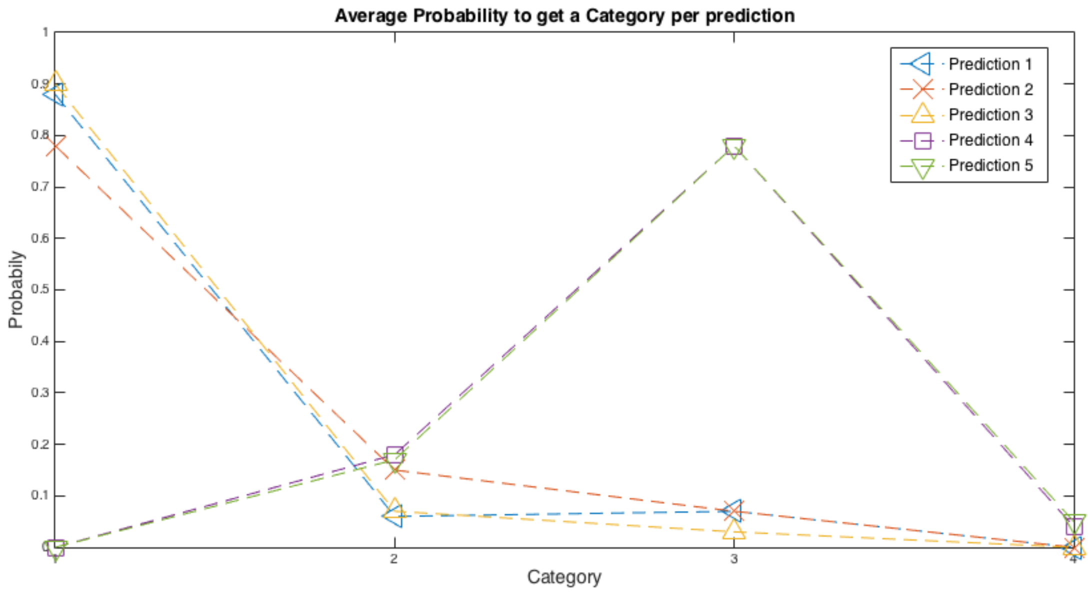


Figure 28 Average probability to get Categories 1-4 for the 5 Predictions

By comparing the probability to get a certain Category in the different 5 scenarios, there is a clear tendency or pattern that can be identified. For the predictions that have a Sea Level Rise estimation of 0 (Prediction 1,2 and 3), the probability to obtain a Category 1 exceeds in all cases the 75%, with a remaining 10% of probability to obtain a Category 2. For the cases where there is a sea level rise of 0,6m, the trend of the graph is different, surpassing, in a 100% of the cases, the probability to obtain a Category 1. In these scenarios, the category with higher probability to occur is Category 3.
What the graph also shows is that Category seems to be in-between both types of predictions; for Predictions with a SLR of 0, most of the measurements have a probability to be under Category 2, while for the scenarios of SLR equal to 0,6, most of the points have a higher probability to be higher than Category 2.

This is important because it differs from what it was seen when discussing the predicted categories. In the prediction analysis, a gradual increase from category 1 to category 2 for more severe scenarios was observed, moving from small water depths to higher ones. However, what Figure 28 shows is that the expected probability is smaller than for category 1. The reason behind this behaviour can be explained as a consequence of the number of variables and information used for the logistic regression.
In the predictions, some points had a unique probability to reach a certain category, such as the points that have been analysed in Table 6. However, the probability, which should be close to 1 appears to be smaller in some points.

In any case, the multinomial regression gives more probability of occurrence to Category 1 (for less dramatic scenarios) and Category 3 for adverse predictions.

7.5 Sensitivity Analysis

In previous parts of this study, several predictions have been carried out, by changing the sea level rise and the return period conditions. However, all the values used are considered "expected values", since there are widely variable and can vary a lot depending on future scenarios. Give this difficulty to forecast the specific values of sea level rise and return period and due to this high uncertainty, a sensitivity analysis has been carried out.

This part aims to understand how does the prediction of hazard evolve under gradual changes of the independent variables. Understanding the trend of the model prediction can be very useful for planning purposes of the government and for the companies located close to the sea, since it would provide a general picture of how the coastline would evolve under certain future scenarios.

7.5.1 Changes in Sea Level Rise

Firstly, the sensitivity analysis has looked at how does the category prediction changes for a low wave return period of 10 years.

For this, the changes in Sea Level Rise have been 9 measurements of 0,05; 0, 10; 0,15; 0,20; 0,25; 0,30; 0,35; 0,40; 0,50. Those conditions were applied to the 3632 points considered for study. The results are presented in Figure 29.

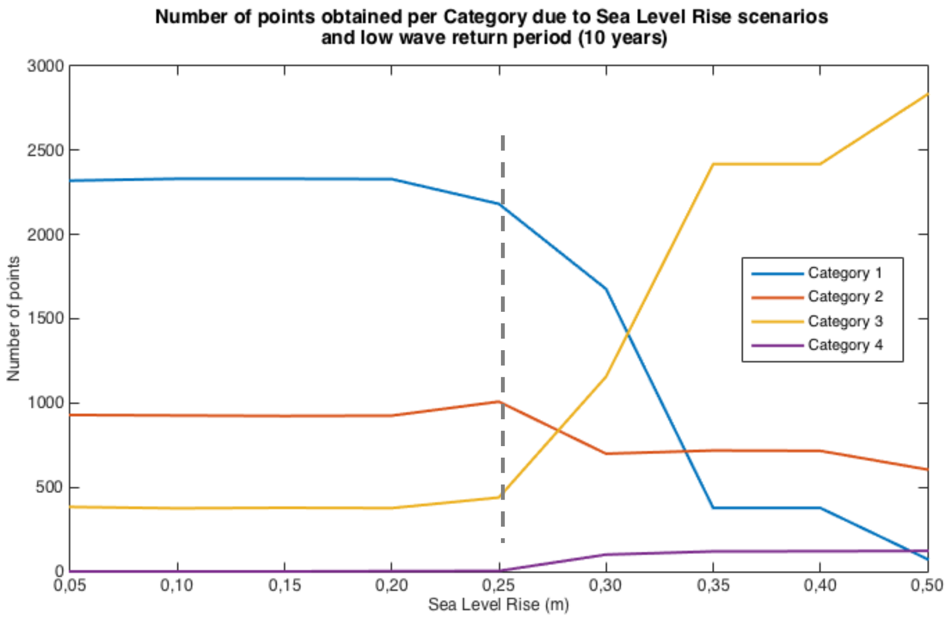


Figure 29 Number of points with category 1, 2, 3 and 4 under a low wave return period of 10 years for changes in Sea Level rise [0,05-0,50 m]

The graph shows the evolution in the number of points for each category as the sea level rise is increased from 0,05 m to 0,5m. As it can be seen, for small values of Sea Level rise, the major Category is 1, with over 2250 points. This represents nearly 65% of the total points measured. For small values of sea level rise, between 0,05 to 0,20, there is no big influence in the total increase of category. The level of the water depth will be mostly between 0 and 0,4 meters (Category 1).

From 0,25m of Sea Level rise onwards, the number of points per category is no longer constant. Points in category 1 drop drastically, while there is an increase in category 2 and 3 points. The biggest rate of increase occurs for category 3 where 717 more points of this category are registered from 0,25 to 0,30. This increase occurs due to a reduction in points of category 1 and also category 2.

It is not until the sea level rise reaches 0,25m that category 3 becomes dominant besides category 2. When more than 0,30m is reached, Category 3 becomes dominant throughout the rest of the measurements.
For 0,35m of SLR, category 4, 3 and 2 points account for almost 90% of the total points measured and only some localised points are at category 1.
For high values of sea level rise, the number of points in category 2 remains constant whereas the number of points category 1 continue to reduce considerably while category 3 points increase. This reinforces the hypothesis that for small increases in sea level rise, the effect on the overall category is very strong.

The graph represents that there is no gradual increase from category 1 to category 2 and category 3 as sea level increases, but rather a jump from category 1 to category 3. The big contribution of sea level rise when predicting category of a point explains that even a small change in sea level rise can shift from category 1 directly to category 3.
For points in Category 4, an increase take place from 0,25 m of sea level rise. After that level is reached, the number of points in Category 4 continue to increase, although for levels of sea level rise above 0,30 m, the rate at which that happens is slower, with nearly the same number of points in category 4 for sea level rise from 0,30 to 0,50m.

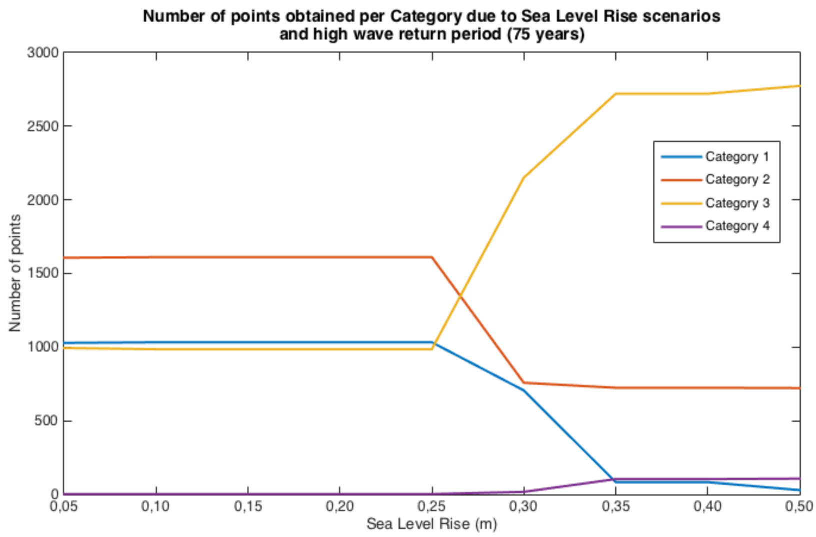


Figure 30 Number of points with category 1, 2, 3 and 4 under a low wave return period of 10 years for changes in Sea Level rise [0,05-0,50 m]

The similar plot has been made with the results obtained for high return period of 75 years (

Figure 30). The main difference from the low return period scenario is that the biggest amount of points in the first prediction of 0,05 m of sea level rise is Category 2 with over 1500 points followed by Category 1 which has similar amounts of points than Category

3. For the first values of sea level rise, and similarly to the previous example, the changes in sea level rise don't have an impact in the amount of point registered for the four categories.

Although there are differences in the number of points obtained per category, some trends and similarities arise by comparing both plots:

- The number of points that register a Category 1 are sharply reduced for increases in sea level rise going from 0,25 to 0,35m;
- Changes in the number of points under Category 2 are mostly reduced by changes in sea level rise from 0,25 to 0,30m;
- Similar to Category 1, Category 3 suffers the biggest variation of the number of points for the range of sea level rise from 0,25 to 0,35m.

For the creation of the graphs, the lowest and the biggest wave return period were plotted. However, despite being different wave conditions, both graphs show the same evolution and similar trend during the same intervals of sea level rise. Therefore, it can be concluded that sea level rise is the dominant variable when predicting the flooding category of the study area, since great variations in wave return period don't seem to affect the evolution of the categories.

Another conclusion drawn from studying the graphs is that the difference in the points obtained for values of sea level rise on 0,25m is similar to the results that would be obtained under a sea level scenario of only 0,05m.

As it can be seen in Figure 29, for low wave return periods, category 1 is the main one whereas for

Figure 30 there is a category 2 dominance. This means that for lower Sea Level Rise, the return period is what determines the major category and the number of points that will be over category one. Analogously, the effect that sea level rise has on predicting the category is only relevant for values of more than 0,25m.

Another way to interpret this behaviour is by looking at the main category for all the scenarios calculated, as shown in Table 9:

		Sea Level Rise (m)								
		0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,5
Return Period	10	1	1	1	1	1	1	3	3	3
	15	1	1	1	1	1	3	3	3	3
	20	1	1	1	1	3	3	3	3	3
	25	1	1	1	3	3	3	3	3	3
	50	2	2	2	3	3	3	3	3	3
	75	2	2	2	2	2	3	3	3	3

Table 9 Main category for the predictions under changes in sea level rise and return period

The table shows that for small values of Sea Level Rise (0,05 - 0,15m), the category gradually shifts from category 1 to category 2. However, once sea level increases over a certain threshold close to 0,25m – 0,30m, the category level moves directly to category 3 without transitioning through category 2. This means that for high increases in sea level rise, there is a big change in the flooding depth to move from 0-0,4m values (category 1) to 0,6-1m (category 3).

7.5.2 Changes in Return Period

The same procedure that has been done for changes in Sea Level Rise is now followed for changes in the wave return period, measuring the prediction of the category with values of Return Period of 10, 15, 20, 25, 50 and 75. For the analysis, two graphs are created, one under a low sea level rise scenario of 0,10m (Figure 31) and another one with a high sea level rise of 0,50m (Figure 32).

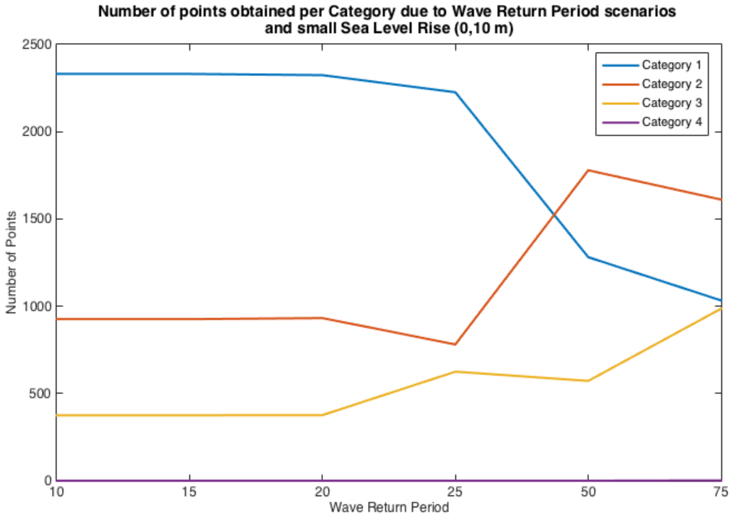


Figure 31 Number of points with category 1, 2, 3 and 4 under a low Sea Level Rise of 0,10m for changes in the Wave Return Period [10-75]

For Figure 31, it can be seen that until a return period of 25 years is reached, the variability in the category stays nearly constant. For values larger than 25, the number of points in category 2 increases, exceeding the amount of points for category 1 in ~40-45 return period. In this scenario, and opposed to what the graphs in sea level changes highlighted, category 2 is the one dominant for high values of return period. Category 3 increases but can not exceed the 1000 points, being below the final number of points obtained for category 1.

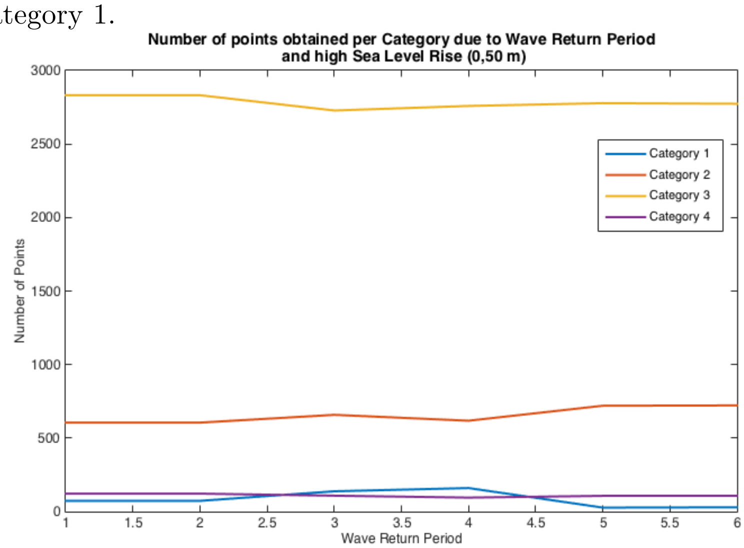


Figure 32 Number of points with category 1, 2, 3 and 4 under a low Sea Level Rise of 0,50m for changes in the Wave Return Period [10-75]

Alternatively, for the case where there is a high sea level rise, the development in the number of points per category has a very different outcome. The first important observation is that there is a wide gap between the number of points measured in category 3 compared to the other categories. This leads to more than 2500 points in category 3 and close to 600 for category 2, leaving a small fraction of the remaining points for category 1 and 4. However, the strongest characteristic in this graph is that it does not seem to be dependent on the wave return period. For different measurements of return period, the number of points per category remains very similar, nearly constant throughout the measurements.

7.6 Overall discussion

Following the study of how the categories are influenced by sea level rise and the wave return period, it has been proved that the study area is more vulnerable to changes in sea level rise.
This has been after analysing the trend in the evolution of the number of categories obtained for changes in sea level rise for a low and high wave scenario is similar, that show that a low or high return period does not interfere in the evolution of he predicted categories for a range of sea level rise (Figure 29 and
Figure 30).
On the other side, it has also been observed that when the values that are observed are the changes in the wave return period, the trend is no longer visible, as changing the values of waver return period with a small and big Sea Level Rise has different effects in the evolution of the categories (Figure 31 and Figure 32).

7.7 Tipping Points and recommendations

Results from the sensitivity analysis have given an indication not only of which are the variables that interfere more in predicting a certain category, but also some reference values of sea level rise and return period that have a high impact in the outcome of the category. Those range of values can be very useful for management purposes, as they become an indication of potential thresholds that could be introduced when dealing with future coastal management measures.

For this reason, it has been considered useful to estimate and establish tipping points, or hydrodynamic conditions that would lead to the category shifting that has been observed in the sensitivity analysis.

- Sea level rise threshold: As shown in Figure 29 and
- Figure 30, most of the changes in the number of points per category is experienced for values larger than 0,25 m of sea level rise. It is for this reason that it has been chosen as the threshold for which relevant management measures should be taken into consideration. Reaching higher values of sea level rise would mean that category 3 events are occurring in most parts of the study area. That would have catastrophic consequences for the beach and all the adjacent services offered.

- Wave Return Period: Even though sea level rise is the main variable when predicting coastal flooding, the return period is still important for values of sea level rise of less than 0,25m (Table 9).
  It for this reason that is has been deemed important to include also a certain range of values for which the return period would shift the category from 1 to 2, under small sea level rise values. Therefore, the value that has been decided to use is a return period larger than 25. It has been chose considering that for values of small sea level rise, the turning point that causes an increase in category 2 points occurs after reaching a return period of 25 (Figure 31).

With the set of tipping values defined, it is now time to measure the potential time when reaching those values would occur. In order to do that, the projected regional rise in sea level up to 2100 in the North-Western Mediterranean Coast has been used, extracted from Sánchez-arcilla et al., 2016.(Sánchez-Arcilla et al., 2016)

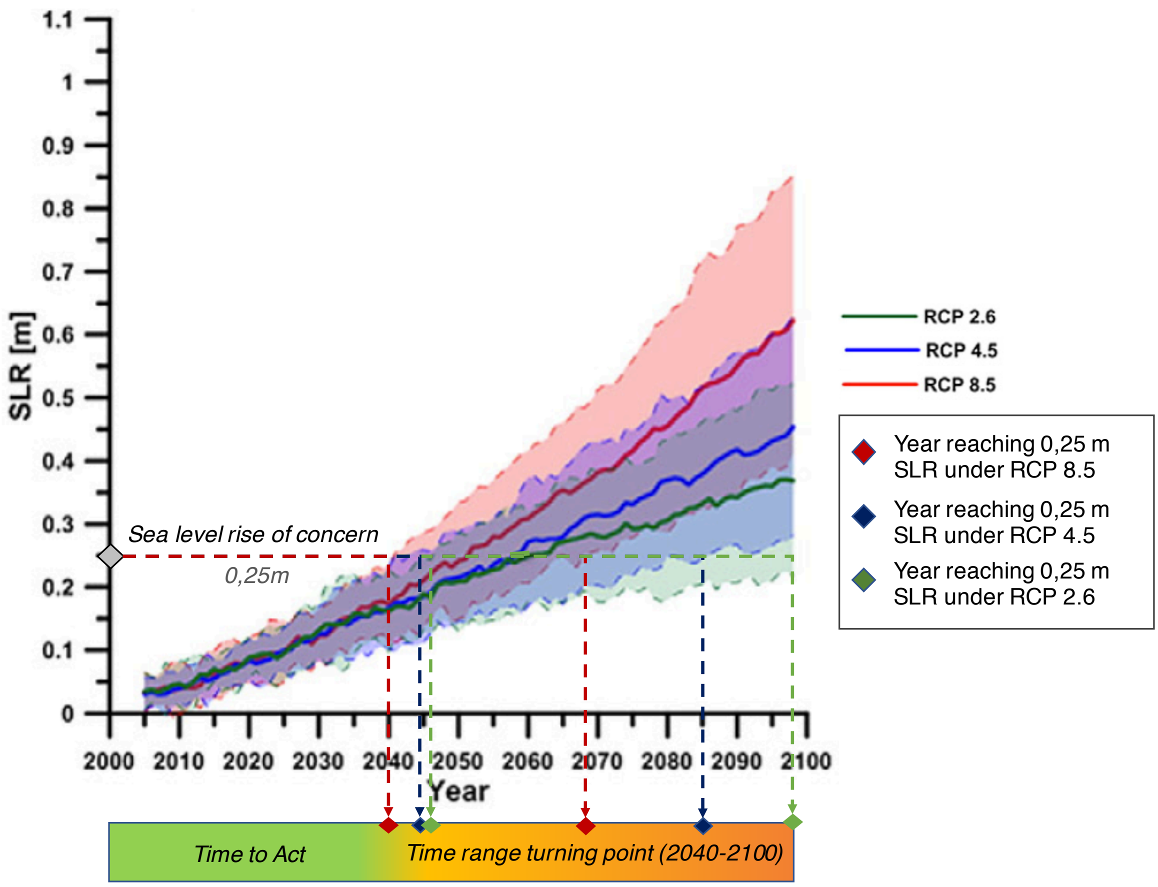


Figure 33 Projected regional rise in sea level up to 2100 at the nearest node of the NW Mediterranean Coast with the threshold values of Sea Level Rise chosen. Source: Own elaboration adapted from Sánchez-arcilla et al., 2016

As it can be seen, under the most severe climate scenario (RCP 8.5), reaching 0,25 meters of sea level rise would occur between 2040 and 2070, in just 23 years. For the softer scenarios (RCP 4.5 and 2.6), a sea level rise of 0,25 m would start occurring only 5 years and 10 years in delay from 2040.

This highlights that the thread of reaching a tipping point is high and that is not that far away from the present day. As it has been seen during the project, reaching a value of sea level rise of 0,25 would not just mean an increase from category 1 to more category 3 events but also confirms that, reaching that point would open a period where categories are widely affected by variations in sea level rise, meaning that small increases from 0,25m onwards would cause more severe consequences than for smaller values, since previous graphs already show that the constant behaviour of the categories only occur for values of sea level rise smaller than 0,25m.

8. Model evaluation and validation

One important aspect after a logistic model is built and the coefficients are calculated is to check whether the model is good and accurate. How well does the model fit the data? Assessing which predictors are the most important and how would the model predict the categories with less or more information is a relevant issue to deal with.
Normally, the first step when building the model is to split the data into two sets: one for training that will be used to fit the model and a testing one that will be applied to the model to check and compare the obtained results.
However, and as it has been seen in Chapter 4, due to the small amount of information available, this separation of the data was not possible to be carried out (since only 6 values were used per point to construct the model).

Another added difficulty was the fact that some points of study had only one category for the 6 measurements, which excluded them for study (since no logistic regression can be done with only one outcome).
These complications resulted in a reduction from 3928 points of study to 3632, leaving out of study information that could have been used for the predictions.
Also, when calculating the probabilities of the prediction scenarios, there were significant points that were clustered as complete or quasi-complete separation, a phenomenon which made it difficult to interpret how would coastal flooding be analysed in those points under uncertain future scenarios.

In order to account for those points that were excluded from the analysis, and also as an alternative way to check whether the model based in points is actually useful for predicting category of flooding, it has been decided to perform the same logistic regression model but for a larger area.

8.1 Polygon distribution

The idea behind this is to divide the whole study area into smaller polygons and instead of performing the logistic regression for every point, and cluster it for each small polygon (and the information of the associated points that are inside the polygon). This way, the logistic regression is not carried out by only a single point with 6 measurements, but within a group of nearby points.

8.1.1 Objectives:

There are several reasons behind proceeding with this validation:

- First, the aim to include the 296 points that were excluded when building the node-based model;
- To analyse whether adding more information to a model has any big effects into the overall prediction of the categories.
- To check how building a model with polygons (and group of points) differs from a model based on specific point measurements.
- To discuss the accuracy of different type of models for predicting coastal flooding. Is it necessary to build a model that creates a logistic regression for each point of study or a more general outcome of the predicted category would not overestimate or exclude relevant information? In other words, is the model information based on a point by point basis more or less detailed and for estimating coastal flooding?
- To check whether for large study areas a more general model based on polygons can still provide accurate information of the potential category of water depth under future scenarios.
- Finally, the model seeks to analyse whether the points that are now included in the model would significantly alter the results of the previous model and to check whether this polygon-based model would suffer also from complete or quasi-complete separation.

8.1.2 Model building

The procedure followed to build the polygon model consisted of:

1. Creation of a *Polygon Grid* in QGIS with the same area of 25x25 over the extent of the study area
2. Intersect the polygon grids with the flooding area for 2015 and a return period of 10 years. The result of intersection results in 700 polygons that cover all the area where there is data available.
3. Once the polygons are created, a *Joint Attributes by location* is carried out, resulting in a *Point layer* where, for every point, there is the associated polygon where the point is (Figure 34).

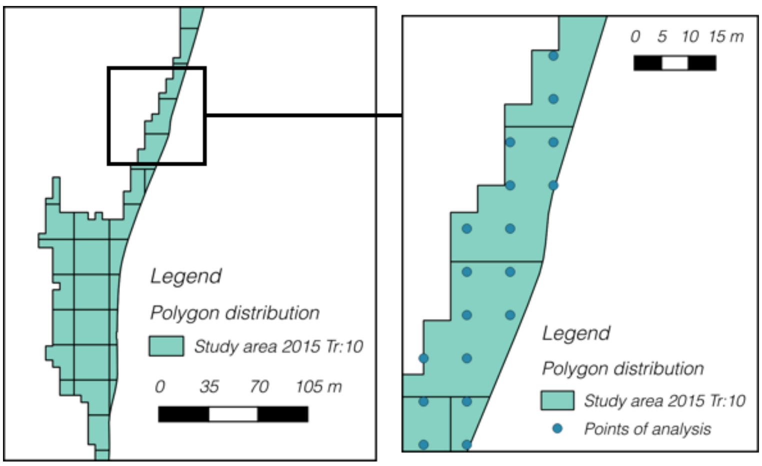


Figure 34 Polygon distribution and detail of points inside each polygon

8.2 Results of the Model

With the 700 polygons of study, the logistic regression model has been carried out in the same way that was done for the points-model. Out of the 700 polygons, the model was able to be computed for 697 of them, leaving only 3 polygons without possible calculation (due to the same problem that was identified for the node-based model, where the category of all the points of study is the same and, therefore, unable to fit a logistic regression).
This result highlights that the number of data excluded is significantly smaller than in the case where the model was performed by points. For the node-based model, there were 296 points excluded. This is reasonable as the polygons are, in most cases composed by several points, which increases the amount of data used to construct the model.
Having to discard only three polygons leaves a model using a 99,5% of the total available data, (contrasting with the 92% that was reached for the model based on points).

Figure 35 show the predicted categories applied in the area of study. As an example, the representation in the map refers to Prediction 1 and Prediction 5 in Zone 3 and highlights which polygons would be more exposed to flooding.

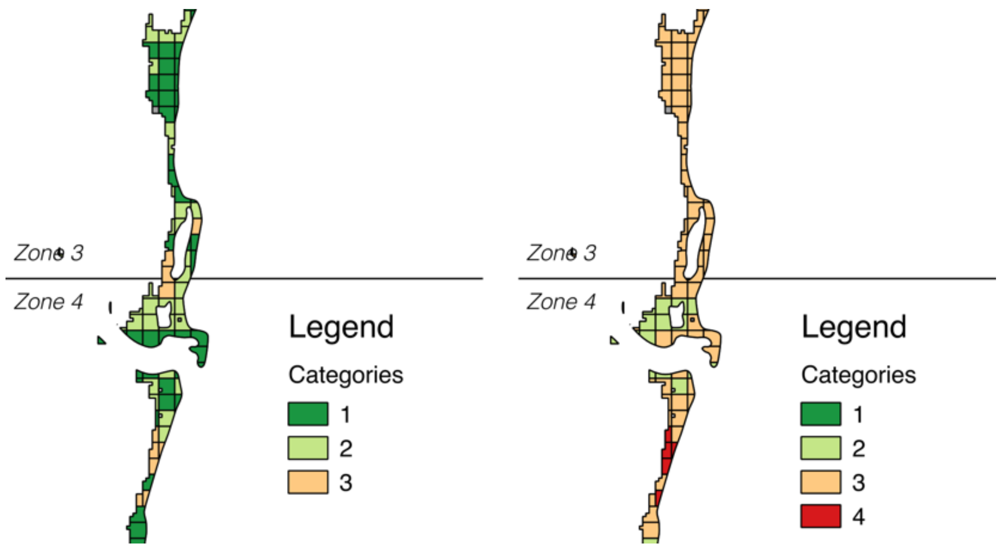


Figure 35 Categories obtained for Zone 3 and 4 for Prediction 1 (left) and Prediction 5 (right)

8.3 Validation of the Node-based model

Now that the polygon-based model has been successfully calculated, the next is to compare the results obtained with the node-based model.
The validation has been made by studying 40 polygons, 5 for each Zone in the study area. An arbitrary selection of 5 polygons per zone was, therefore, selected. The idea is to compare the category that was predicted in the polygon-based model with two values: (1) the mean category of the points inside the polygon for the same Prediction scenario and (2) the category that is most present in the polygon of study. Then compare the results and consider correct if the category predicted for the polygon is the same as the category more present for the points that are inside the mentioned polygon.

This way, it is possible to check whether assuming a Polygon-Based approach would give consistent and similar results to the Node-based model or if the wide range of points inside the polygon would give rather different predictions.

The mean category of the polygon has also been calculated so as to check whether the predicted category of the polygon overestimates or underestimates the outcome category.

The results of this validation have been measured for a low Prediction 1 scenario (with an assumed Sea level rise of 0 and a return period of 10) and for a high Prediction 5 scenario (that assumes a Sea level rise of 0,6 and return period of 100). Both measures can be checked in Table 10 and Table 11.

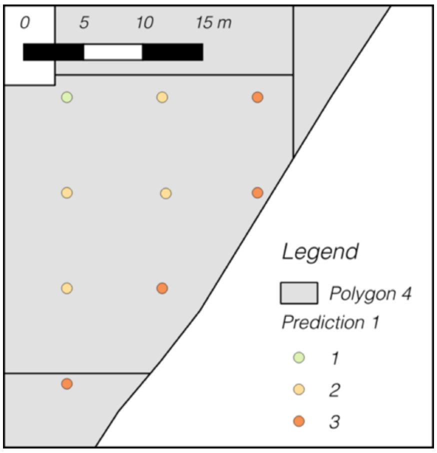


Figure 36 Example of Polygon 4 with the points of study inside

Figure 36 visually shows an exemplification of the procedure followed for validation as it is depicted in Table 10 and Table 11. As it can be seen, polygon 4 has 8 points with different categories predicted in the scenario with no sea level rise and a return period of 10. What has been done is counting the predominant point in the polygon (in this case Category 2) and also the mean which is 2,25. With it, if the major category is equal to the category that was predicted for the polygon, which in this case is also 2, then the point model is considered to be accurate and not very far from the results obtained for the polygon-based mode.

Prediction 1 Scenario (SLR:0; Tr:10)

Polygons	Mean Point Prediction	Max. Frequency category	Cat. Prediction of Polygon	Validation
ZONE 1				
4	2,25	2	2	
13	1,6	1	1	
26	1,4	1	1	
31	1,75	2	2	
43	2	3	1	-2

ZONE 2			
54	1,5	1	1
69	2	2	2
72	1,85	1	1
77	1,6	2	2
86	2	1	1
ZONE 3			
92	1,75	1	1
627	1,3	1	1
112	1,6	2	2
638	1,3	1	1
124	1,4	1	1
ZONE 4			
163	2	2	2
182	1,5	1	1
198	2,5	3	3
654	1,4	1	1
220	2,16	2	2
ZONE 5			
661	1,2	1	1
669	1,5	1	1
267	1,11	1	1
276	1,7	1	2
288	1,3	1	1
ZONE 6			
302	1,2	1	1
315	1,6	1	1
328	1,7	1	1
348	1,5	2	2
382	1,6	1	2
ZONE 7			
397	1,6	1	1
410	1,5	1	1
429	1,2	1	1
440	1,4	1	1
475	1,5	1	1

ZONE 8				
496	1,5	1	1	
516	1	1	1	
531	1,4	1	1	
708	2,05	2	2	
581	1,5	1	1	

Table 10 Comparison between polygon and point based model for Prediction 1

Out of the 40 polygons measured, 92,5% of them predicted the same Category as the predominant category that the points inside the polygon have. This means that a polygon-based model for Prediction 1 represents a similar behaviour as the individual node-based model.
The range of categories in this project takes values of 1, 2, 3 or 4. However, calculating the average category from the points inside the polygons results in values that are in-between. Analysing them gives the result that 29 out of the 37 (78%) polygons overestimate the mean value of the category whereas 8 underestimate it. This has no special importance when analysing the prediction of a certain category, but as it will be discussed later, it does have an impact when measuring probabilities.

Prediction 5 Scenario (SLR:0,6; Tr:100)

Polygons	Mean Point Predictions	Max. Frequency category	Prediction of Polygon	Validation
ZONE 1				
4	3,5	4	3	
13	3	3	3	
26	3,2	3	3	
31	3,3	3	3	
43	3	3	3	
ZONE 2				
54	2,9	3	3	
69	3	3	3	
72	3	3	3	
77	3	3	3	
86	3	3	3	
ZONE 3				
92	3	3	3	
627	3	3	3	
112	3	3	3	
638	3	3	3	
124	3	3	3	

ZONE 4			
163	2,14	2	2
182	2,5	3	2
198	3,5	3	4
654	3	3	3
220	3,6	4	4
ZONE 5			
661	2,8	3	3
669	2,8	3	3
267	3	3	3
276	3,2	3	3
288	3	3	3
ZONE 6			
302	3	3	3
315	3	3	3
328	3	3	3
348	3	3	3
382	3,1	3	3
ZONE 7			
397	3	3	3
410	3	3	3
429	2,6	3	3
440	2,7	3	3
475	2,5	3	2
ZONE 8			
496	2,8	3	3
516	2,5	3	3
531	2,7	3	3
708	3,2	3	3
581	2,75	3	3

Table 11 Comparison between polygon and point based model for Prediction 5

The validation for the Prediction 5 scenario is also remarkably high, with 36 out of the 40 polygons valid. This proves that for changes in sea level rise and return period, the model based on polygon gives similar or quasi-identical results as the point based model.

Different to what it was observed in Prediction 1, only 6 polygons out of 36 polygons overestimate the mean value of the category. The behaviour in this case is opposite to prediction 1, where there were more polygons overestimating the mean category of the points inside the polygons rather than underestimating.

8.4 Study of the expected probabilities

After computing the node-based regression model, it was seen that there was a great amount of points that suffered from complete or quasi-complete separation (as show in Chapter 7). This itself did not pose any problem for the prediction of the expected category but it did have an impact on the expected probability, since a complete or quasi-complete separation means that the logistic model is not used to determine the category. This caused some problems when analysing how would the probability change for each point, since in all points that had a complete and quasi-complete separation, the R code gave a 0 probability.

For this reason, analysing how the polygon-based model gives the probability of reaching a certain category is very important.
In order to get the expected probability to reach Categories 1, 2, 3 and 4 for the 5 Prediction scenarios was carried out through "R" following the same procedure as for the node-based model.
After the calculation, the probability distribution of categories for Prediction 1 to 5 have been obtained. A small part of the results obtained from Prediction 1 can be seen in Table 12.

Polygon ID	**Category Predicted**	**Probability**			
		Category 1	Category 2	Category 3	Category 4
5	3	0,344666247	0,266205942	0,389099429	2,83817E-05
6	3	0,499777895	0	0,50011871	0,000103395
7	2	0,333314258	0,388424892	0,278257661	3,18929E-06
8	2	0,399967018	0,413902766	0,186130217	0
9	1	0,397388658	0,373478393	0,229132949	0

Table 12 Extracted table of the category probabilities and categories obtained for Prediction 1 (SLR:0; Tr:10)

Opposed to what it was analysed for the node-based scenario, none of the polygons of study suffered from complete or quasi-complete separation.
In addition, as it is shown in Table 12, the probability of reaching a certain category is more variable and gradual, with very small differences between given probabilities. For examples polygon 5, has a probability of having a Category 3 of 0,38 whereas reaching a Category 1 is very close with 0,34. This means that, even though the predicted category is 3, the probability to have Category 1 is still very important.
Does this mean that the model is biased, or that the probabilities of having a certain category are too close that the model is easily alterable?
If the only thing that is evaluated from this model is the predicted category, then the model could not provide as accurate results as the node-based model. That is because the node-based model performed the logistic regression for each point, whereas this polygon-based scenario considers several points when determining the probability.
However, if the predicted category is contrasted with the probability distribution, then this model provides very useful information on how an area will behave under certain conditions.

For the node-based scenario, the probabilities did not help reflect how were other categories, out of the predicted one, represented in the model. The lack of enough measurements caused the probability to be nearly 0 for all the categories that were not the predicted one. This provided a concise value of the potential prediction under a certain Prediction scenario, but made it impossible to suggest how would the other categories evolve during those scenarios. There was no information on how points with a certain Category were susceptible or not to reach a Category 3.

With a polygon-based model, the range of probabilities that are obtained are larger and more distributed, which gives a better estimate on how susceptible to change in the near future.

Therefore, a polygon-based model not only can assess a category (which in the previous section of this project it was validated to be accurate compared with the node-based polygon), but it can also explain in general terms how each zone can evolve under the different scenarios.

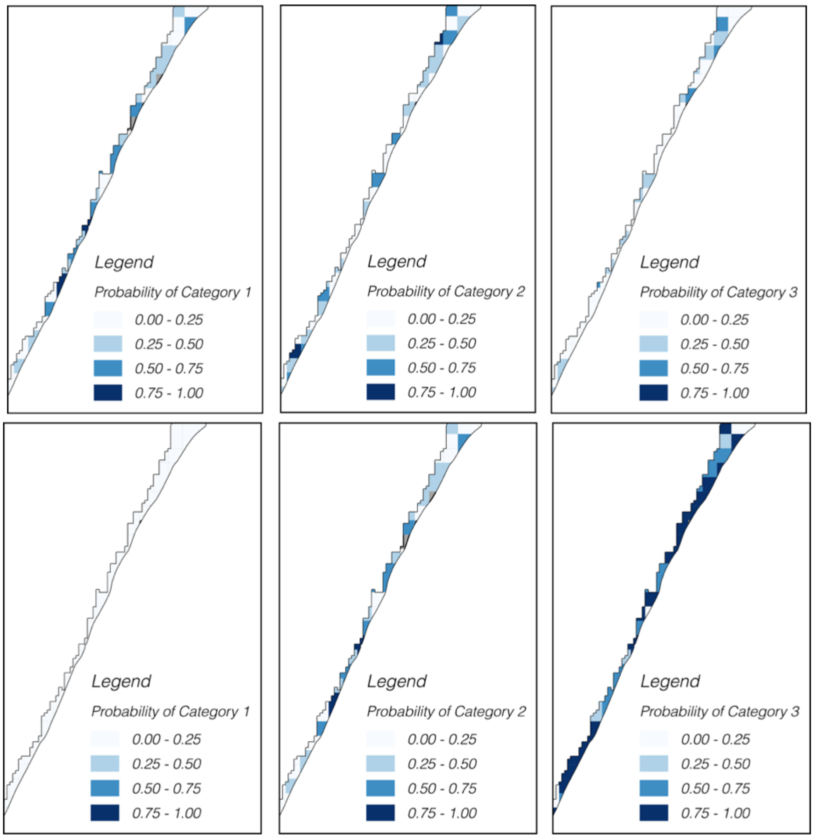


Figure 37 Comparison of expected probabilities for Category 1, 2 and 3 conditions for Prediction 1 (top) and Prediction 5 (bottom) in the North Region.

Figure 37, Figure 38 and Figure 39 are taken as an example of how the probabilities of the different categories could be used to track the evolution of the water depth for the different regions. This is very beneficial because, not just gives information on the category that each polygon will have but also what is the probability of other categories to occur.

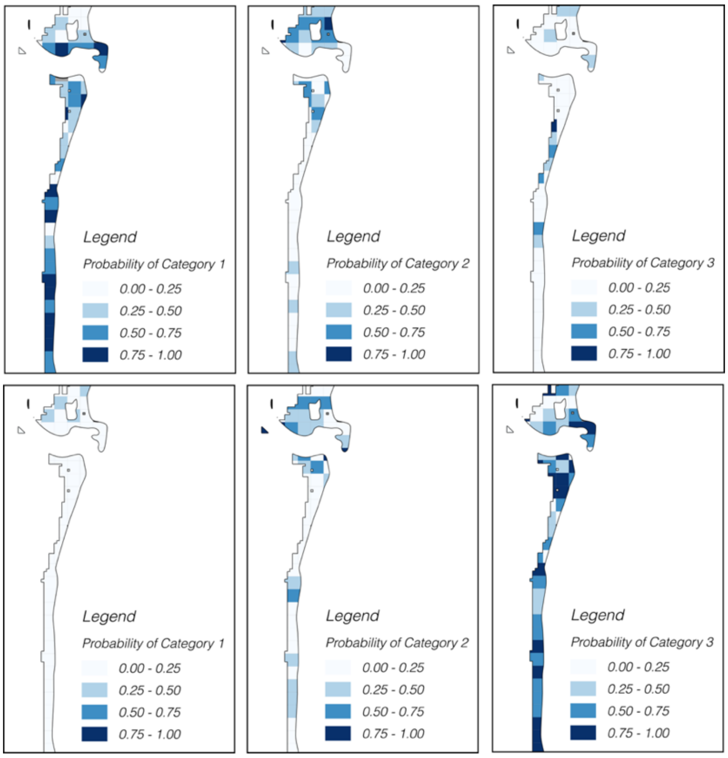


Figure 38 Comparison of expected probabilities for Category 1, 2 and 3 conditions for Prediction 1 (top) and Prediction 5 (bottom) in the Centre Region.

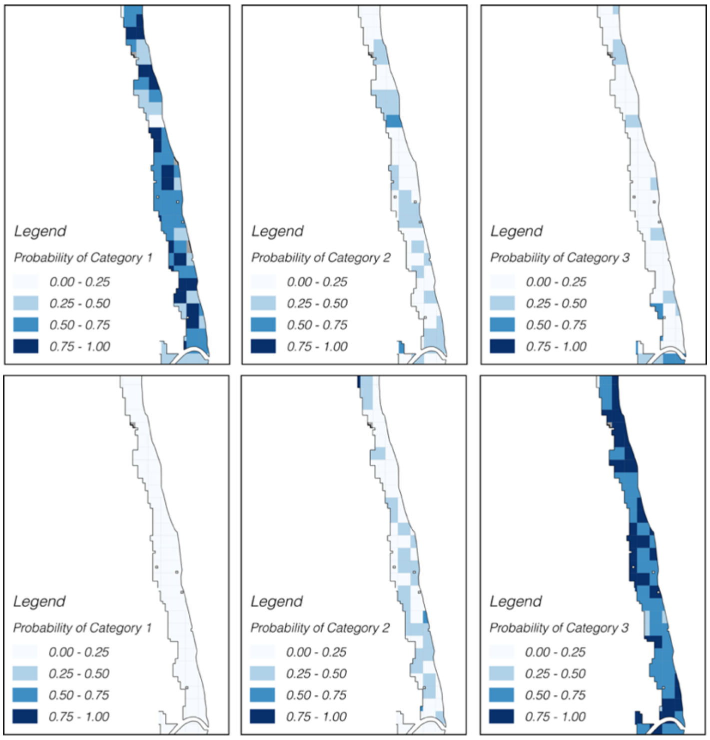


Figure 39 Comparison of expected probabilities for Category 1, 2 and 3 conditions for Prediction 1 (top) and Prediction 5 (bottom) in the South Region.

8.4.1 Comparison of node-based and polygon-based regression model

After analysing the results obtained for the polygon-based model and by checking its validity and resemblance to the model that was built through nodes, the only remaining task is to compare both models in order to get the strengths and weaknesses. Table 13 highlights both models and its main characteristics.

As a general comment, both models provide an accurate representation of the predicted categories under different scenarios, with different probability distribution and spatial coverage as the main difference.

Node-based Model	**Comparison with regard to**	*Polygon-Based Model*
The node-based model is based on 3632 points of study.	**Initial Dimension of the model**	The polygon-based model consists of 700 polygons that contain 3928 points
296 points were excluded because the logistic regression was not able to fit the logistic regression. It represented a 7,5% of the total.	**Errors**	Only 3 polygons had to be excluded for analysis because, for the input data, all had the same category, so the logistic regression was not able to be carried out.
For most of the points, there is only one probability (with a very high value of 0,99) whereas for other categories there are probabilities of 10^{-5}. For the initial Predictions (1,2 and 3), 12% of the points suffer complete or quasi complete separation.For larger Prediction scenarios (4 and 5), the number of cases is reduced to 0,5%.	**Probability of Categories**	There is a distribution of probabilities with similar values for different categories (they go from the range of 0,20 to 0,70). There are no polygons that suffer from complete or quasi-complete separation so all the 700 polygons manage to give a distribution of probabilities for each category.
The node-based model provides a point per point prediction with a high probability of occurrence.	**Accuracy of the predictions**	The polygon categories can overestimate or underestimate the average category of the points inside. However, they have the same category as the largest category represented by the points inside.
It gives a predicted category for every point, so there is a wider spatial coverage of the effects of sea level rise and return period. An additional positive aspect is that the results obtained are similar as the ones that are extracted through the polygon-based model which means that	**Strength**	The probability distribution accounts for a wider spectrum of possible categories, due to the fact that the polygon is built with the measurements of several points. Having a larger spatial representation can help interpret

with less data and computational cost, an accurate prediction is obtained.		how different regions of the study area are affected by SLR and return period.
A local analysis of the category ends with less data to build the model which can lead to more extreme results. Even with the predicted categories obtained, the probability distribution is very biased, challenging the interpretation of the changes in category for future scenarios.	**Weakness**	The polygon distribution has been established randomly by defining a cell size thought the study area. For planning and management purposes, specific regions should be identified and delimited.

Table 13 Comparison of the node-based and polygon-based model

9. Spatial Analysis and management application

9.1 Study area flooding

So far, this project has worked in the creation of a model to predict the evolution of the Category of flooding. For the creation and analysis of the model, two scenarios of 0 sea level rise and 0,6 m level were taken together with wave return periods of 10, 50 and 100. However, the analysis was done taking the measurements delimited by the flooded area of the less aggressive scenario (top left picture in Figure 40).

Working in a delimited region was necessary to determine the accuracy of the model and to analyse tipping points and the range of values of Sea Level rise and return period that would affect more the beach of study.

However, there is a great amount of information that was excluded from the analysis, mostly the flooded area for more extreme scenarios. As Table 14 shows, there is whole amount of information available about the spatial extension of the flooding and the regions that, due to more extreme scenarios, would be prone to flooding. The chapter will deal with these data, in order to focus more on how will the water depth of the study area evolve, determine the main assets at risk and propose the potential measures that could be implemented to prevent those consequences to occur.

	Conditions		Total Number of points
	SLR	Return Period	
2015	0	10	3.928
		50	5.764
		100	9.212
2100	0,6	10	15.593
		50	54.920
		100	64.000

Table 14 Total number of points per layer and the result number of extractions for the first simulation. The reference map defined with blue is the reference taken for the creation of the model.

In Figure 40, the total flooding area is represented for the different scenarios that will be analysed in this chapter, outlining the extension of the water along the coastline and its surroundings.

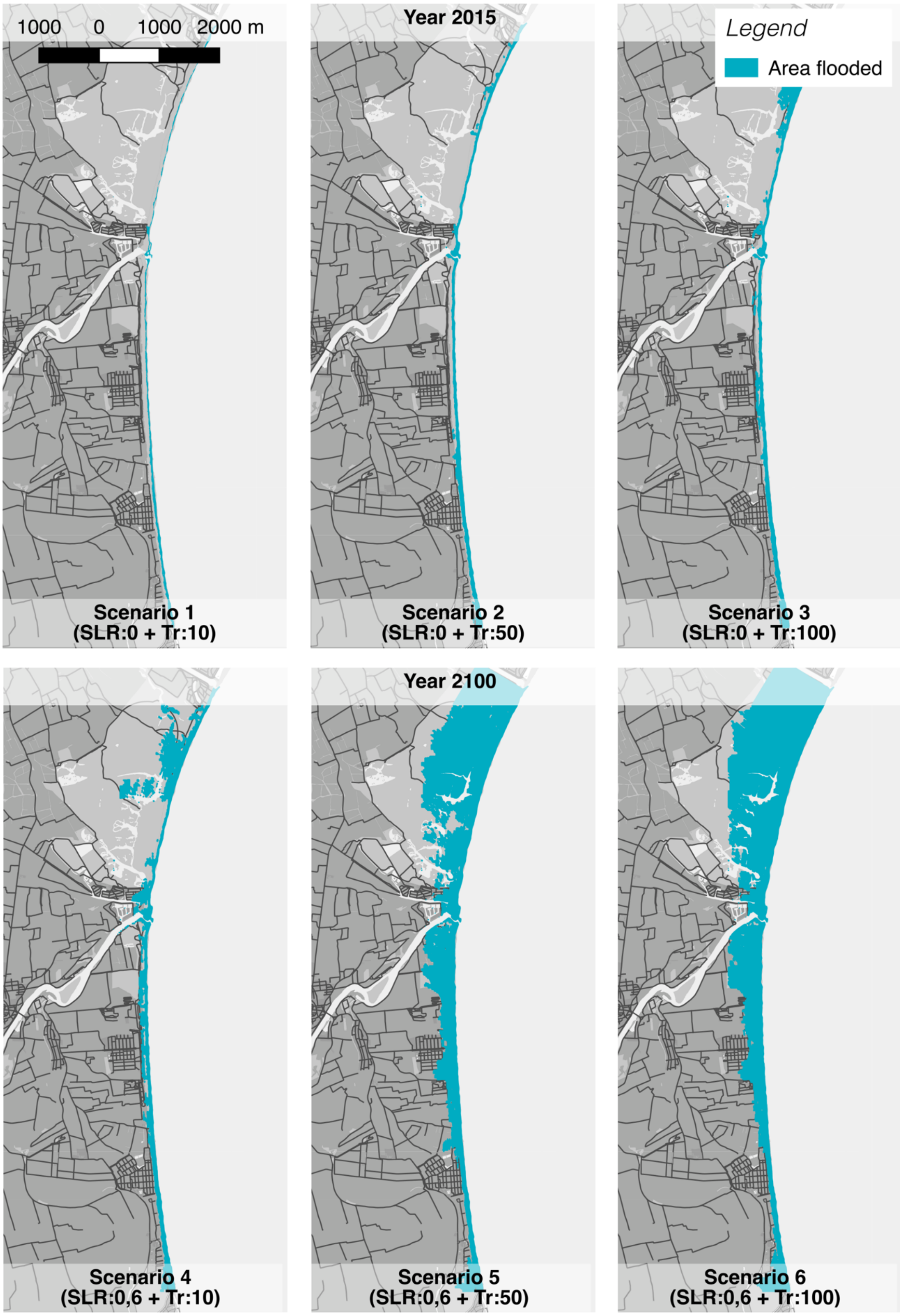


Figure 40 Inundation Area for the 6 scenarios proposed

As it can be seen, for models of no sea level rise contribution with only an increase in the wave return period, the evolution of the flooded area tends to expand along the beach of Sant Pere Pescador. Table 15 also shows the changes in the total flooded area for the different events.

Wave Return Periods	Flooded area (in Km^2) with no sea level rise	Flooded area (in Km^2) with 0,6 m of sea level rise
1 in 10 years	0,255071	0,990107
1 in 50 years	0,372833	3,984852
1 in 100 years	0,589985	4,674653

Table 15 Total Flooded area for various return periods predicted assuming current conditions (2015) and future scenario of 0,6 m sea level rise (2100)

By analysing the spatial distribution of the flooded area, it can already be seen that, similar to what was seen in previous chapters of this project, the Northern Region is more prone to flooding, with a larger extension of water for sea level rise scenarios. In general, even for the less severe scenario, the beaches will suffer from flooding, already in case where wave return periods were considered.

9.2 Objectives

For this purpose, this chapter will aim to:

- Determine the evolution of flooded area for the scenarios where no sea level rise occurs but there is an evolution in the wave return period and for the scenario with a sea level rise of 0,6m and the same changes in wave return period.
- Visualize in 3D the elevation profile of the beach and the adjacent regions in order to observe if there are natural barriers against flooding and to check if there is a terrain characteristic that might explain why the northern region would have larger flooding extension (as it was discussed in the previous chapter).
- Analyse, for the different flooding scenarios, how would the land uses be affected and determine the assets at risk by the increase of flooding.
- Determine potential measures to increase the adaptation to climate change and to reduce the vulnerability of the study area. Provide a list of potential activities that could be included in an Integrated Coastal Management plan.
- Combine the measures for coastal flood management with the tipping points to determine the location and the time of application of the measures.
- Provide a qualitative methodology for the analysis of the risk, based on the assets at risk and the different climate scenarios.

9.3 3D Analysis of the flooded area – Dunes as a natural barrier

In previous chapters, it was seen that the distribution and extension of flooding varies depending on the location. Measuring where the biggest amount of points in Category 3 and 4 were located showed that, for the Northern Region of the study area, the number of points in Category 4 doubled the ones in the Centre and Southern Region (Check chapter 7.2, *Representation of predicted categories*)

This trend has also been analysed in this chapter by Figure 40, that shows that not only the Northern region is expected to reach category 4 water depths, but also that it is expected to flood more surface than the other regions. In order to provide an explanation to these differences between regions, a 3D model representation will be done through a Digital Elevation Model together with the water depth obtained from the simulations dataset.

Since the study area is located in a low-lying area, the elevation of the DEM model has a smooth appearance. In order to accentuate the contours and texture of the surface, a terrain analysis has been carried out in QGIS, transforming the existing DEM model into a Hillshade (Figure 41).

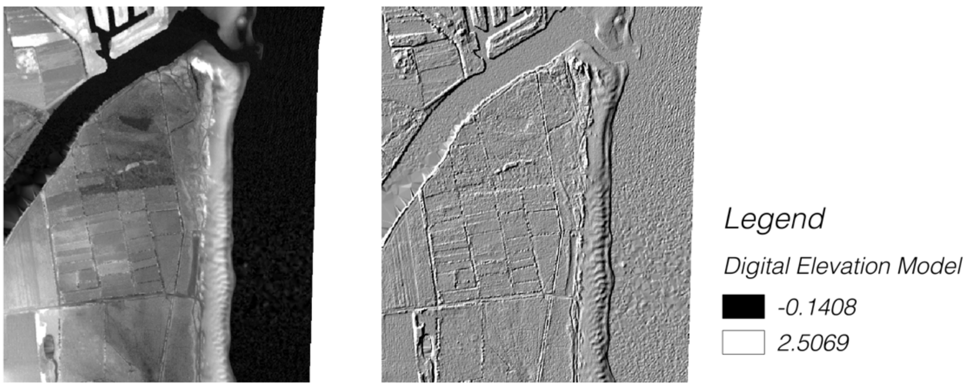


Figure 41 Transformation from the DEM model (left) to a Hillshade (right) applying the Terrain Analysis tool of QGIS

The representation of the topography of the study area with the hill-shading method is based on simulating the effect of natural light on earth's surface. The visual result of the method is formed by the change of the tone of the portrayed surface, which is due to the effect of light and the differentiation of orientation on each point of the surface.

The 3D model has been obtained through ***QGIS 2 three JS,*** a Python plugin available for QGIS that exports terrain data and map canvas images to a web browser.
As Figure 41 shows, the area of study has small elevation differences, reasonable considering that is a low-lying area. In order to stress the terrain characteristics. Therefore, it was deemed appropriate for visual purposes, to give more vertical exaggeration to the model.
The satellite images of the region have been included in the representation of the 3D model as well as a render of the elevation model coloured depending on the value of the raster file.
The overall results are shown in Figure 42.

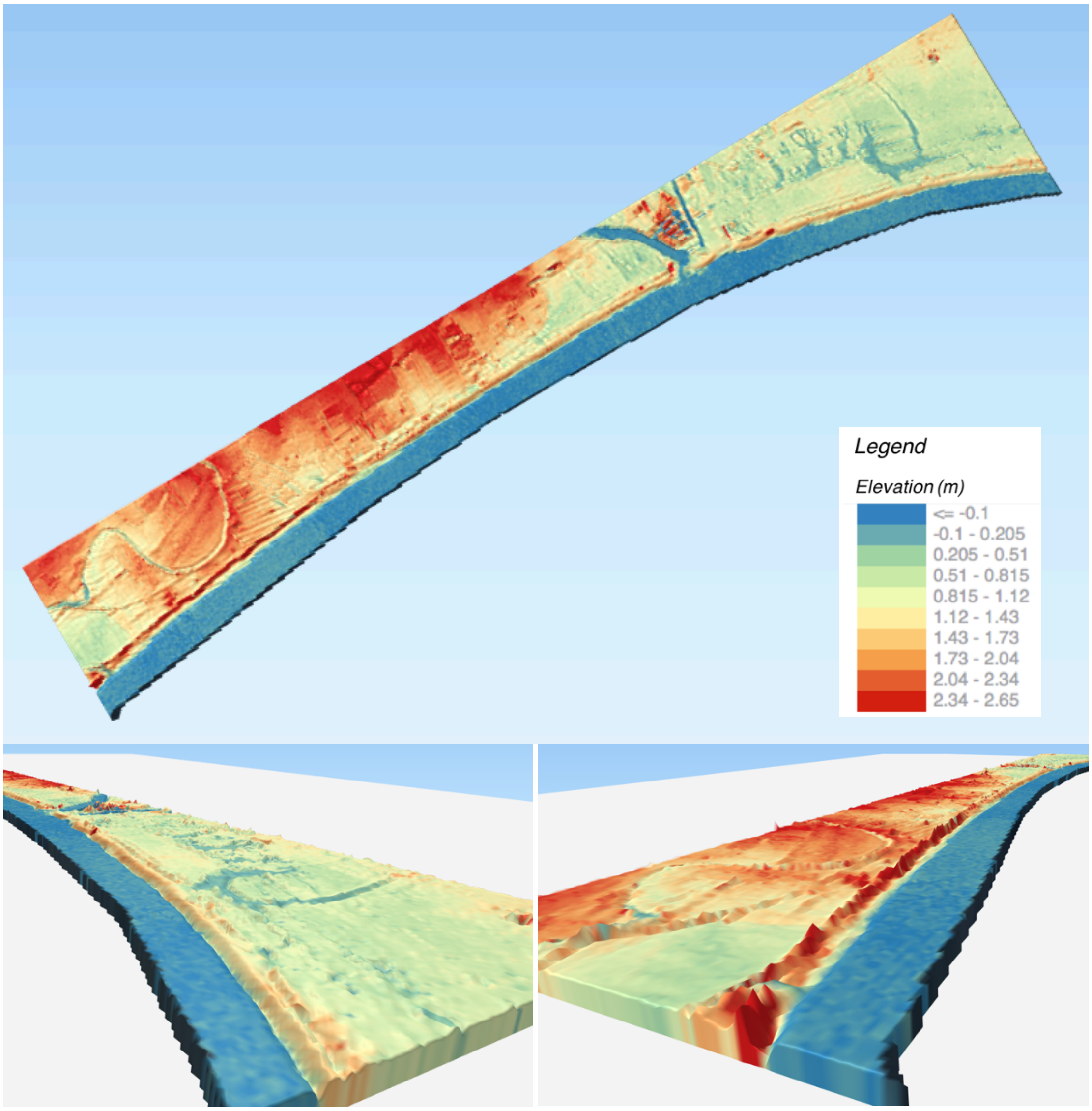


Figure 42 3D Representation of the Digital Elevation Model. The whole study area is represented in the top picture, and closer zoom is made to the norther region (bottom left) and the southern region (bottom right). The surface floor is given an exaggeration parameter to facilitate the interpretation.

As it can be seen, by representing the Digital Elevation Model, the northern region has a more homogeneous floor compared to the southern region. The type of beach is considered to be dunes, so the 3D model shows how the surface roughness can play a key role in preventing coastal flooding to occur.

If the elevation data is taken into consideration, it can also be seen that, for the areas in the south, the elevation takes values between 1,2 and 1,7m whereas for regions in the north the elevation is smaller, approximately 0,5 and ~1m
The natural morphology of the beach, as it has been exemplified in this section, can suggest why it is in the north where larger and more abrupt flood water depth are registered; An information that can also be very useful when deciding flood protection measures.

Similarly, the 3D model can also support the representation and visualization of the flooding extension. As an example, the Scenario 6 has been also characterised.

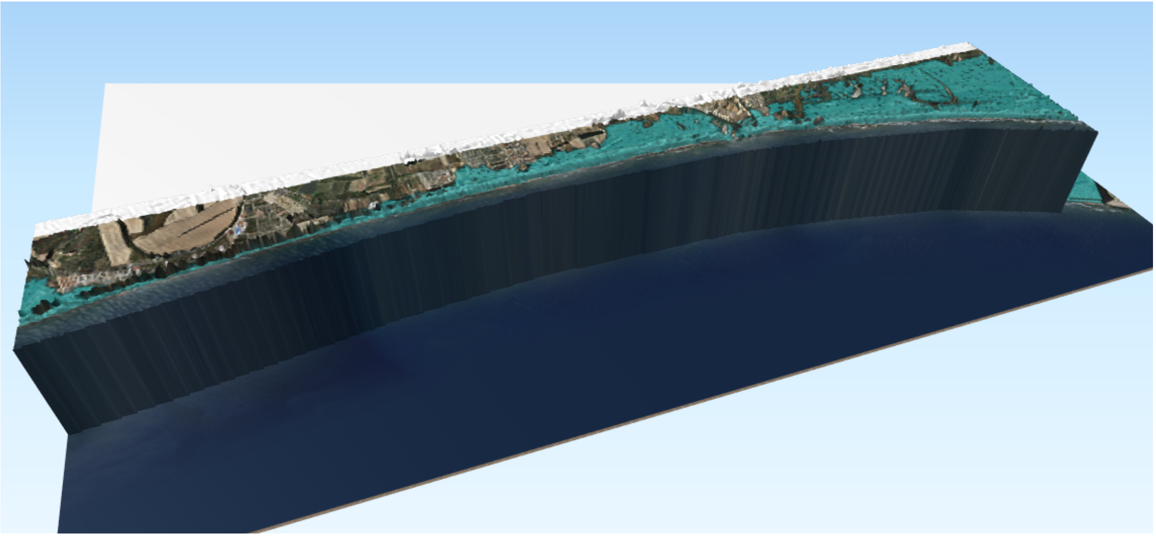


Figure 43 3D representation of the flooding extension under the most sever scenario of Sea Level Rise of 0,6 m and a wave return period of 100 years. The layer of water is coloured in light blue.

9.4 Methodology

Most of the work that has involved spatial treatment of the information has been developed through QGIS 2.18.2.

9.4.1 Data collection

Dataset	Source	Resolution	Description of Data
Raster flood maps	Pinyol et al., 2015	Pixel 8x8	6 Raster maps of the water depth as a consequence of changes in sea level rise (0 and 0,6 m) and the wave return period (10, 50 and 100)
Vector flood maps	Pinyol et al., 2015	-	Transformation of the raster maps into vector layers (polygons).
Land use map	Institut Cartogràfic i Geològic de Catalunya (ICGC)	1:25000 Precision: 2,5m	Set of vector layers that identifies the elements of a region (transport, buildings, water spaces and also elevation).
Ortophoto	Institut Cartogràfic i Geològic de Catalunya (ICGC)	Pixel size: 50x50 cm (High resolution)	Aerial photograph of the Sant Pere Pescador beach and the surroundings susceptible to flood.
Digital Elevation Model	Institut Cartogràfic i Geològic de Catalunya (ICGC)	Regular mesh 2x2 m	A 3D representation of a terrain's surface created from terrain elevation data

Table 16 Total Flooded area for various return periods predicted assuming current conditions (2015) and future scenario of 0,6 m sea level rise (2100)

9.4.1 Procedure

There are several ways to work towards adaptation measures. What is important to bear in mind, when starting, is the available information and its details. It has been decided to carry out a methodology based on a Vulnerability Reduction Approach. Figure 44 shows the different steps that will be followed to conclude with future considerations and adaptation measures.

The following methodology is based on existing procedures combined with personal suggestions of what should be included in the analysis.

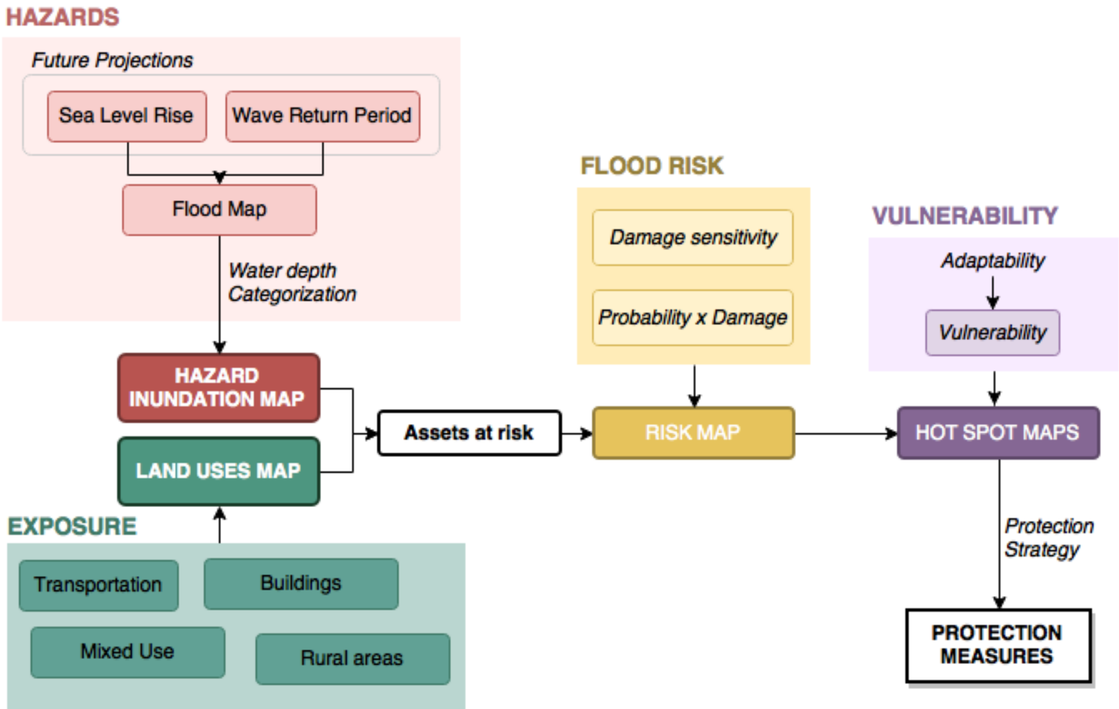


Figure 44 Flowchart showing the general methodology with the main components (hazard, exposure, flood risk and vulnerability). Source: Own elaboration adapted from *(Muis, S, 2015)*

9.5 Determining the Assets at risk

The first step towards determining the assets at risk is to obtain the Hazard Inundation map and the Land Uses map.

The Hazard was already obtained in previous chapters of the project. It is characterised by the water depths measured in the area for the different scenarios.

9.5.1 Land Use maps

The Land use map describes the different assets that will be analysed. It is the representation of where in the study area can public buildings, vegetation or water resources be found.

This map can also include elevation points as well as detailed information on the general slope. For the region of study in particular, it was checked that the slope was smaller than 0,5%.

Figure 45 shows the zoning map with the land uses that will be analysed.

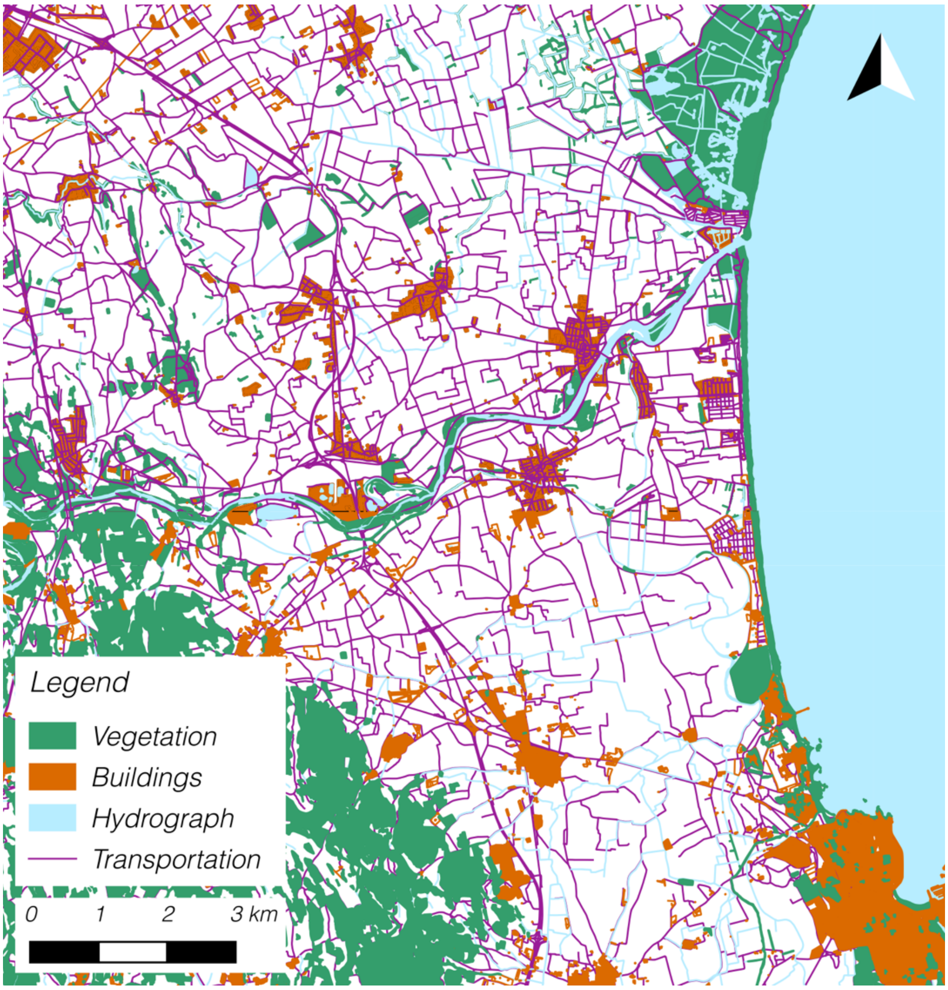


Figure 45 Land uses map delimited for the study area with the 4 types of land uses that will be analysed: Buildings (private and public), Water Resources, Transportation roads and Vegetation and Land cover (wetlands, beach and vegetation).

9.5.2 Transportation facilities

During a flood event, it is essential to maintain communications open and make roads and railways accessible. This is why it is very important to measure which would be the main transportation infrastructure that would be affected from flooding.

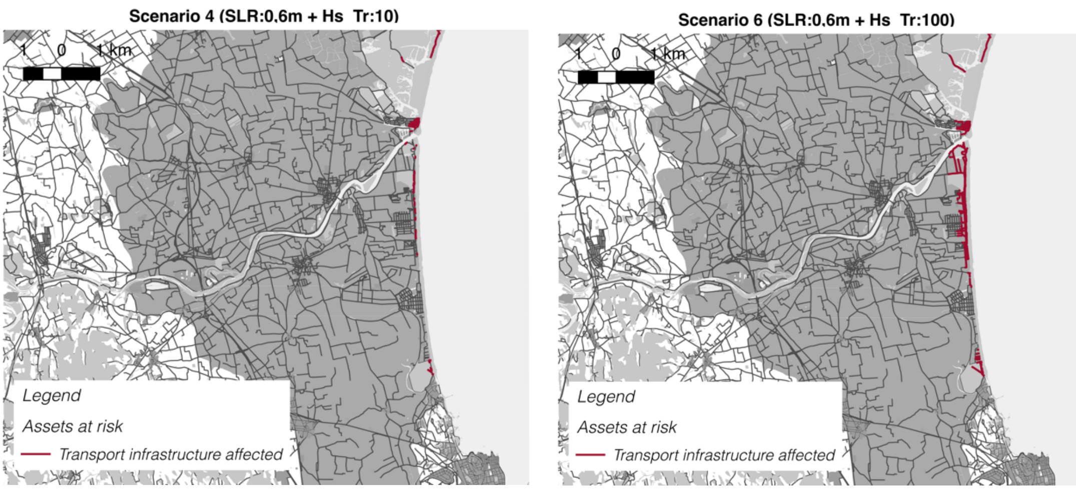


Figure 46 Transportation routes that would be affected under Scenario 4 and 6

In order to calculate the Km of roads affected, the vectorised layer of the area flooded under the 6 scenarios has been intersected with the transportation line layer of the land uses vector map. Figure 46 shows two examples of the kilometres of affected roads for the Scenario 4 and 6. In addition, the total length of affected transport infrastructure for each prediction, can be represented, as in Figure 47.

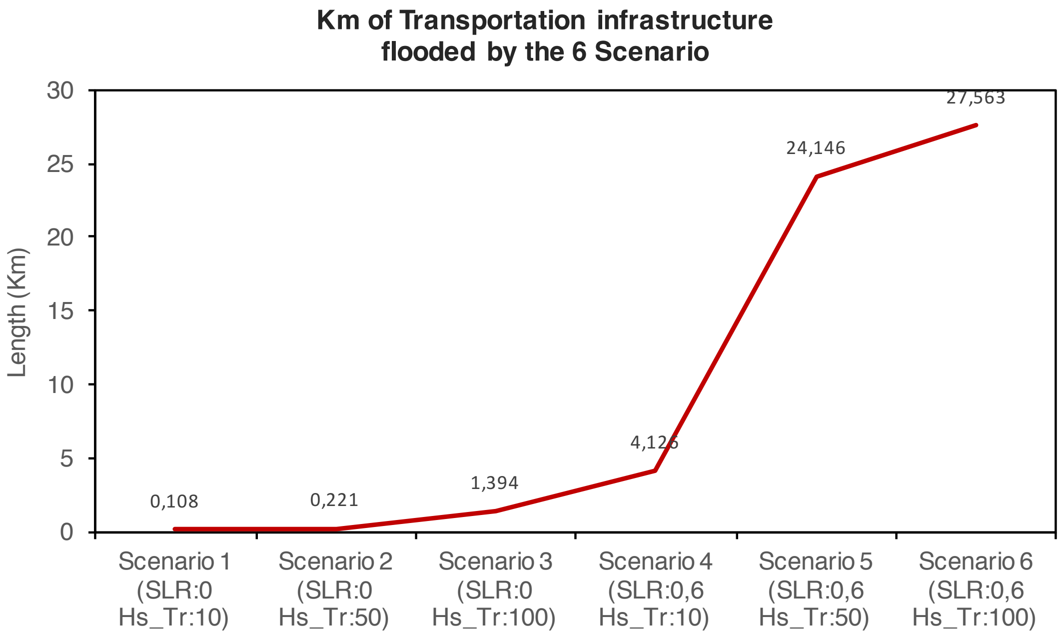


Figure 47 Km of transport infrastructure that would be affected for the 6 scenarios

Results show that for Predictions with no sea level rise (not very likely to occur), only more than 1 km of roads would end up being affected with a high wave return period. However, if predictions are met and the 0,6m sea level rise is reached, the study area could account for more than 20 Km of roads flooding.

In QGIS, checking the attribute table of the transportation layer map, allows to observe which would be the type of roads affected. In general, there are 4 types of roads that would suffer from flooding:

- VIA103 Low traffic urban road not paved nor coated.
- VIA109 Non-urban unpaved path
- VIA113 Urban path (without detailed information)
- VIA 091 Non-urban paved road

As it can be seen, the four types of road are characterised for being low-traffic and mostly non-paved surfaces. It is the type of paths and narrow roads that connect the beach with the adjacent infrastructure. A positive aspect to consider is that not highway or railway would be affected by flooding, at least for the scenarios measured. Another potential benefit is that, since it is considered to be non-paved urban roads, they are a type of roads not likely to connect with major big transportation roads, which, in case of flooding, would not cause that much disruption as in the case of a highway.

9.5.3 Vegetation and Land cover

There are three covers that would be affected by flooding: The Beach, Wetlands and vegetation. The habitats that would be more disturbed would be the wetlands and the beaches. As Figure 48 shows, wetlands will suffer the largest flooding, followed by the beach.

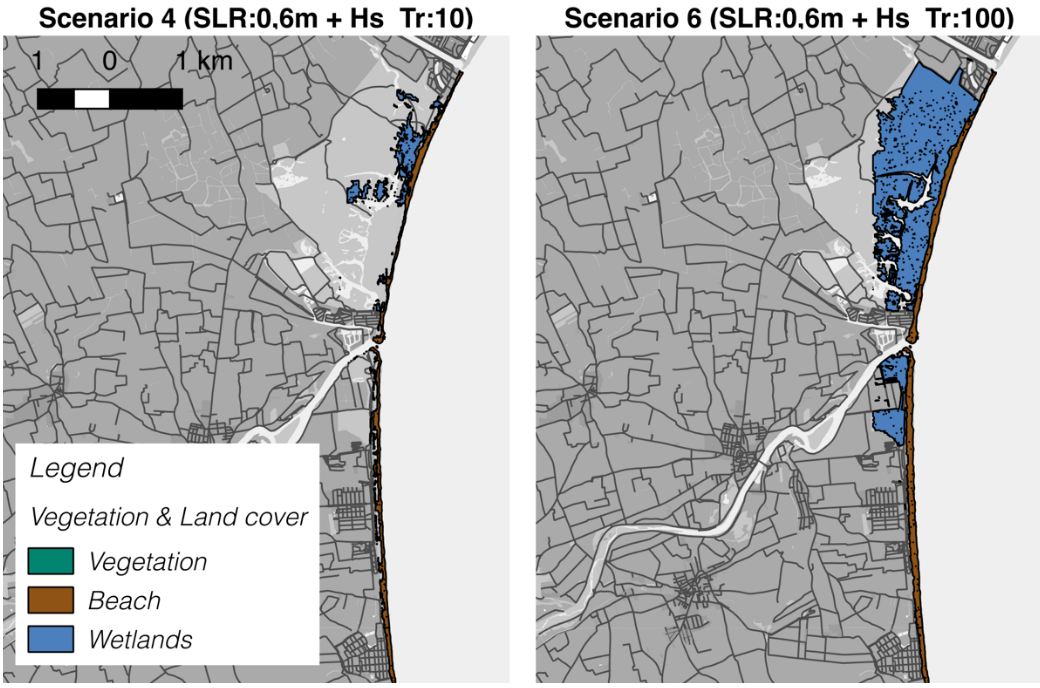


Figure 48 Area of Land and Vegetation cover that would be affected for the 6 scenarios

To illustrate this extent with numbers, it can be seen that for scenarios of 0 level rise but an increase in the wave return period, the kilometres of land and vegetation would reach the 6km. If sea level rise is accounted, the value would drastically increase reaching for the Scenario 6, a total 22,543Km of land and vegetation flooded. If the area is measured, it can also be shown that small predictions of no sea level rise (Scenario 1-3), the area that would be affected will increase from 0,2 Km^2 to nearly 1 Km^2. In the worst-case scenario, 3,58 Km^2 of land would be affected.

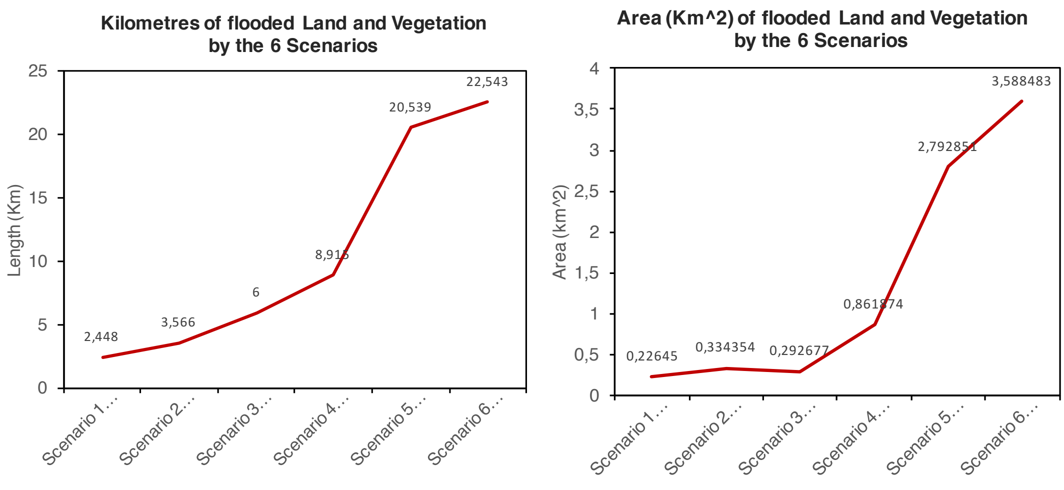


Figure 49 Km and Area of Land & Vegetation cover that would be affected for the 6 scenarios

As it has been mentioned, flooding will affect the beach and, to a larger extent, wetlands. This would have a direct impact into a natural protected habitat, the Natural Wetland Park of l'Empordà. Currently, they are officially declared national natural interests due to its integral zoological and botanical reserves. The denomination of an “integral natural reserve” is considered the highest level of natural protection in the Catalan legislation. The map on Figure 48 clearly shows that a vast area would be affected from flooding. Indeed, out of the huge extension of the wetlands, the area that would be flooded is the area under the label of “integral natural reserve”.

For what concerns the beach, and as it has been explained in detail in previous chapters, the morphology and dune configuration would give more protection in the southern areas so it would be insufficient for the northern region to avoid severe flooding.

9.5.4 Water resources affected

This section includes all the water bodies that would also suffer from sea water. Figure 50 represents the affectations also for two chosen scenarios.

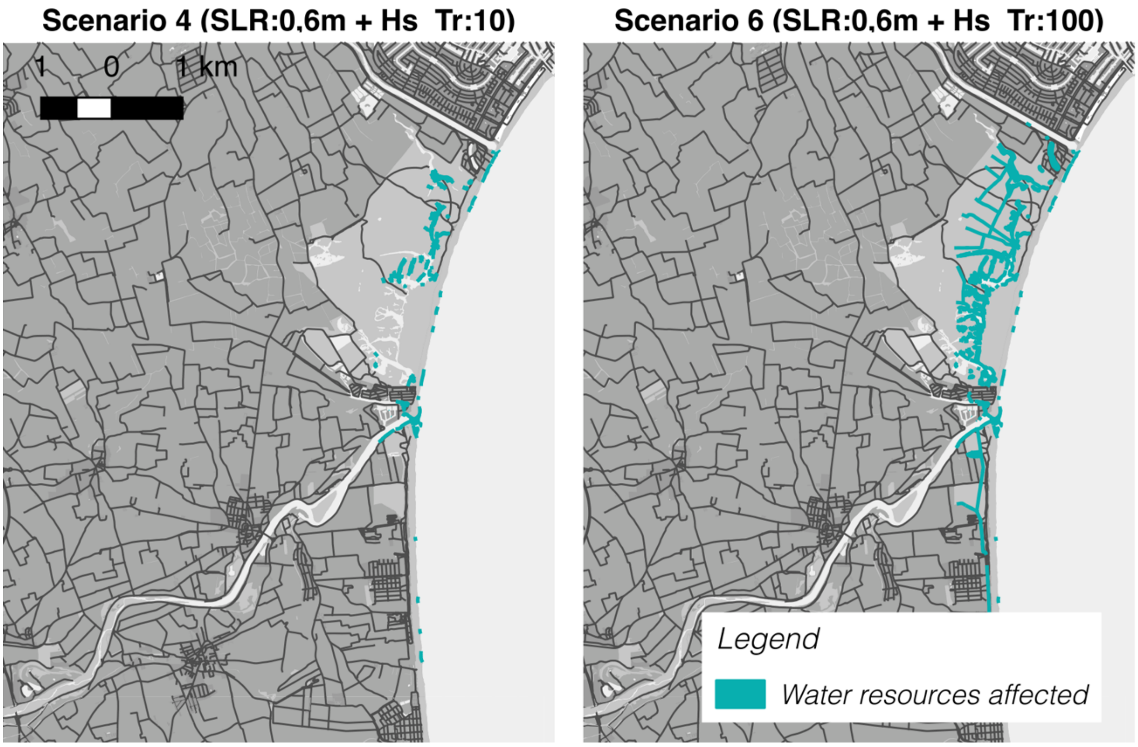


Figure 50 Extension of water bodies that would be affected for the scenario 4 and 6.

This section is especially relevant because it presents one aspect that has not been much discussed during this project and is the water quality. Water bodies located inside the land can have different characteristics of chloride concentration, the water acidity or the water pollution concentration.

Instead of analysing these water bodies affected from flooding from a risk analysis perspective, a special consideration should be made to the connection of the water resources that would be affected, since sea water could be transported in other regions outside from the area shown in the Figure 51.

It is also worth comparing the increase in the kilometres and area that will be flooded. As Figure 51 shows, only in the first Scenario, more than 2 kilometres would be flooded. Accounting for the worst-case scenario, the kilometres of water that would be affected increases to nearly 30Km. This is a huge amount of water affected.

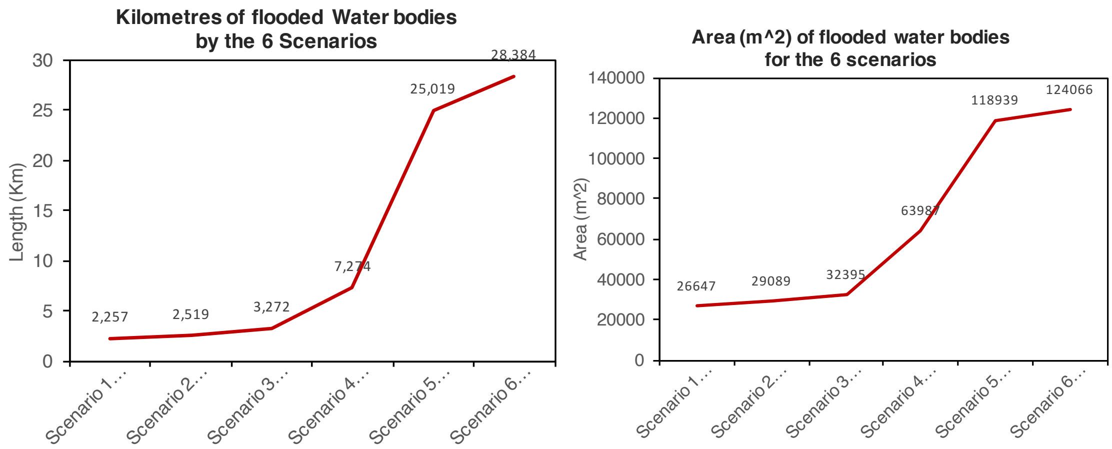


Figure 51 Km and Area of Water Bodies that would be affected for the 6 scenarios

There are three types of water bodies that would be affected. The most relevant one would be parts a river watercourse, water enclosures (a small proportion of the total) and also lagoons and lakes.

9.5.5 Building infrastructure affected

The Sant Pere Pescador beach has several businesses near the coast, since it is a very important tourist attraction. From restaurants, Hotels, a sailing school, open picnic areas and bars...there are several outside activities that would be affected by flooding (Figure 52).

Taking a closer look at the buildings and services that would be affected, it has been seen that there are 4 areas specially at risk. Figure 52 provides a detail picture of the affected zones:

- Area A: 2 camping and several leisure activities. It would affect also a local surf shop.
- Area B: One of the biggest camping is located, with adjacent restaurants and shops. It has also two bars located closer to the beach.
- Area C: It is where the sailing school is located (coloured in purple). Next to it there are two motorhome camping.
- Area D: A limited number of houses that are close to the sea would be affected, not only because their proximity to the sea, but because there are surrounded by water.

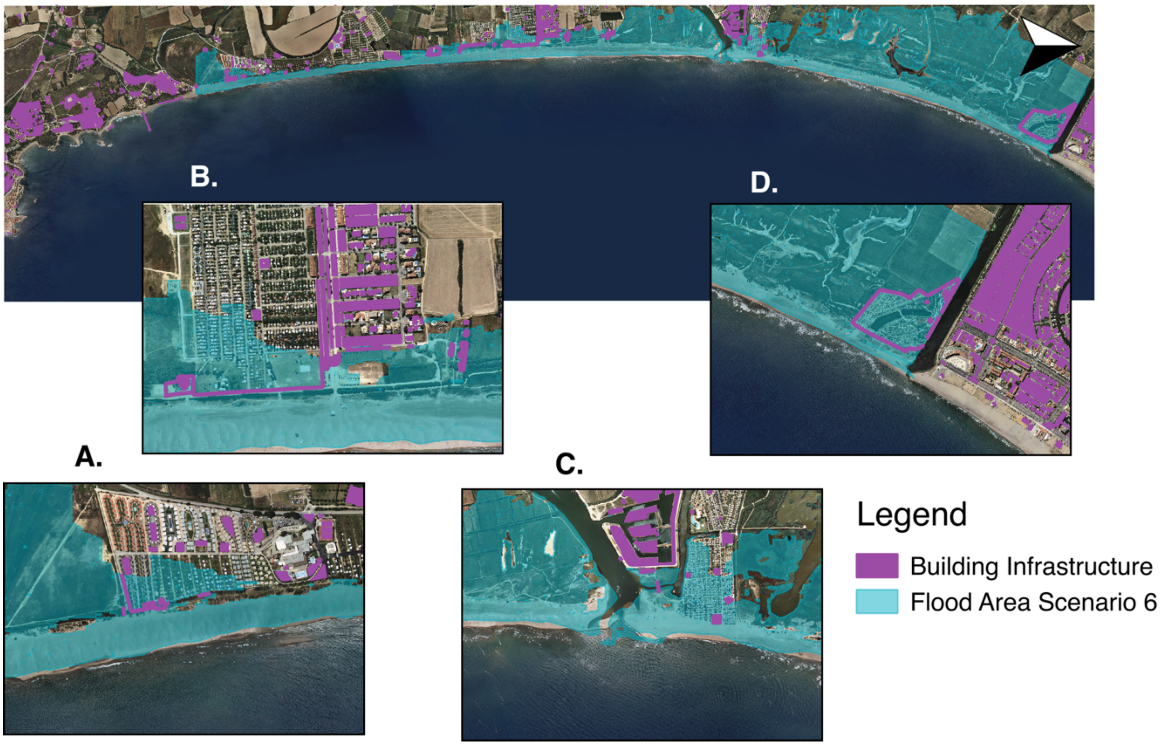


Figure 52 Building Infrastructure representation with the flood area map for a Scenario 6

The most serious aspect to consider is that there are over 9 Camping areas spread around the area of study. Campings are usually very exposed to pluvial flooding since they tend to be big flat surfaces with tents and motorhomes. In addition, an increase in the sea level rise could thread the existence of the camping putting at risk the lives of all the visitors and tourists staying there.

9.6 Risk Analysis (Application case for the Scenario 6)

There are big uncertainties in how climate is going to change and, with it, how much is the sea level is expected to rise. The dynamics of the variables involved in coastal flooding widens up the range of scenarios that are likely to occur, making it a complex task for people, companies and governments to prepare in advance.

Building a logistic regression model with information on sea level rise and wave return period to predict coastal flooding has facilitated the understanding on the potential flooding area and the water depth extension for 6 different scenarios.
Overall, even in the scenario where there is no sea level rise accounted, the beach results highly affected and, without any measures taken, by just an increase in the wave return period, a vast extension of the Sant Pere Pescador is expected to flood.
This opens the question on how to deal with these expected affectations, acknowledging that, as this project showed, already by 2040, the sea level rise could contribute to severe flooding categories.

To give answer to all of these questions and to analyse which regions would require first adaptation measures, this last section of the project will develop a risk analysis applied to the most severe Scenario (SLR: 0,6 m and Hs_Tr: 100). To complement this analysis, some recommendations will be given to alleviate the impacts of flooding and to increase the adaptation and resilience of the area.

In order to do the risk analysis, three sources of information obtained during this project will be used:
1) Patterns of flood height under different hydrodynamic scenarios: The models show that for SLR values larger than 0,25m there is a steep increase in the flood category.
2) The 3D Digital Elevation Model of the study area that shows that the Northern region is more vulnerable to flooding than southern region
3) The main assets at risk (transportation, water resources, infrastructure and vegetation).

9.6.1 Methodology of study

The risk analysis has been carried out following the guidelines of the *ISO 31000 of Risk Management – Principles and Guidelines.* It is an international recognised standard that integrates the concept of risk from different perspectives.

ISO 31000 proposes seven elements interconnected that show how a Risk Management procedure should look like:

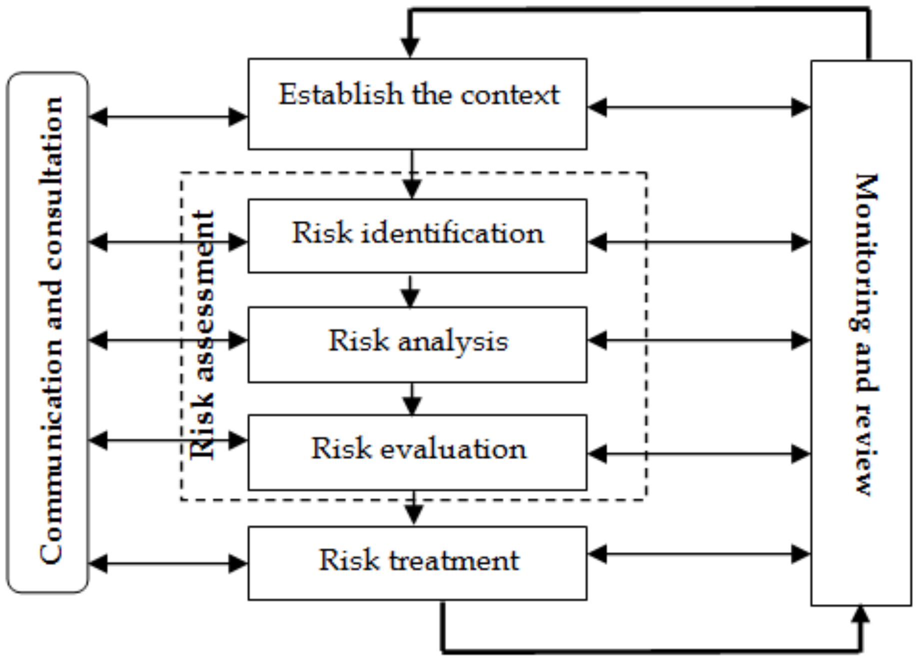


Figure 53 Flowchart of the Risk Management procedures as established by ISO 31000

The first stage consists in establishing the context and analysing the threads, pressures and opportunities of coastal environments. This has already been discussed during this project, concluding that coastal flooding is/and will be highly influenced by an increase in the SLR and wave-return period.

9.6.2 Risk Assessment

The risk assessment is the overall process of risk analysis and risk evaluation. In this part of the risk analysis, the consequences and the probability of occurrence are determined, in order to measure the level of risk:

$$Level\ of\ Risk = Probability\ x\ Consequences$$

Once the risk analysis is carried out, the results can be given in a Risk Matrix, followed by a risk evaluation that will result in the measures to be implemented to accept, avoid, reduce or transfer the risk (Figure 54).

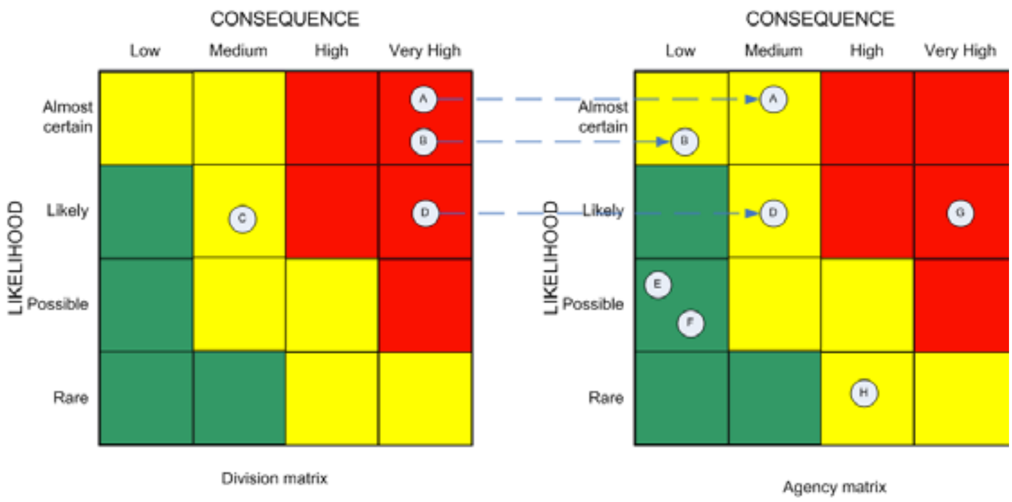


Figure 54 Different levels of risk and the adaptation measure procedure carried out to reduce risk. (NSW Treasury, 2012)

9.6.3 Determining the Consequences

The type of consequences that will affect the Sant Pere Pescador Beach and its surroundings have been already identified in previous chapter, resulting in:

- Transportation consequences
- Water resources consequences
- Vegetation and Land cover consequences
- Buildings and infrastructures consequences

To measure the level of consequence, 4 categories have been defined: “Very High”, “High”, “Medium” and “Low”.

Table 17 shows a proposed categorization of the assets at risk that assigns indicators to determine the consequence level that each asset at risk could suffer. This is done with the aim to classify the assets at risk obtained in the previous chapter with a certain consequence.

Assets at risk	**Consequence Level**			
	Low	Medium	High	Very High
	1	*2*	*3*	*4*
Transportation	-Small paths of rural unpaved areas - Narrow roads outside dense population - Easy drainage and recovery	-Small unpaved urban roads - Stretches of urban roads flooded - Recovery of the affected roads in $<$ than 24 hours	- Secondary roads affected -Urban and rural roads affected - Stretches of main roads and highways affected - Large maintenance required to continue the use - Incapacity to maintain the service in $>$1 days	-Principal roads highly affected (highways, railways...) - Main connection routes for public transport affected - Blockage of transit that difficult maintenance works - Incapacity to maintain the service in $>$2 days or more
Vegetation and Land Cover	-Minimal surface affected - Easy drainage in a few hours - Open spaces & zones allowed to be flooded	- Small damages to land uses and vegetation cover - Recovery and drainage of water in $<$24h - Open flooded areas easily accessible	- Local protected habitat affected - Intrusion of pollutants into protected environment - Threads to flora and fauna of the region	-Impossibility to continue with functions of the land -Protected land affected (National natural resource or local - Flora and fauna at risk -Large extension of land affected for periods $>$ 1 day
Building infrastructure	-Abandoned buildings -Pump and drainage measures easily applicable	-Drainage measures required during 1 day or more	-Institutional buildings and businesses close to the sea -Damage in large number of homes that require the owners to move -Small Affectations in campings	-Disruption of the services -Flooding of campings and dense populated areas -Inability to -Flooding in power supply system buildings and main services (stations, shops,...).

Water resources	-Punctual floodings in closed water bodies -Flooding in end streams of a river	-Flooding in rivers and several river streams -Punctual flooding of open water bodies	-Flooding in drinking water resources -Flooding in principal and secondary rivers -Technical cleaning of water affected by -Intense work to remove salinity excess and coastal organic pollutants from closed water bodies	-High saltwater concentrations intrusions - Intrusions of sea water and sediments in principal rivers -Flooding of protected water environments (wetlands, lakes...)
Flood depth	0-0,4 m	0,4 – 0,6 m	0,6 – 1 m	> 1 m

Table 17 Proposed Consequences table to analyse the risk. Source: Own elaboration

The idea in this stage of a risk analysis is to be as precise and concise as possible. In most risk analysis, this section is complemented with damage curves that include the cost that flooding would have to each of the assets at risk. Given the difficulty to assess the risk from an economic perspective due to the lack of numbers, it has been decided to treat the consequences from a qualitative perspective.

In a risk analysis, and during the discussion between the stakeholders involved in the adaptation measures, this table should be validated by all of them, including the perspective of the local government, the owners of private companies, businesses...

An agreement of the consequence levels is vital to find a common vulnerability reduction approach.

With the given table, now the assets at risk that were previously intersected with the flooding area layer, can now be evaluated to establish which consequence level would they be. In order to do that, the procedure followed is:

1. Assign to each asset at risk, a certain consequence level (Very High, High, Medium or Low), depending on the description given in Table 17. As an example, if the transportation map is taken, for each type of road affected by flooding, a consequence level will be given, depending on whether the road is unpaved, paved, if it is highway...
2. Once each attribute has been categorized for each consequence level, the Assets at risk map (Input Vectorial Layer) is converted into a Raster Layer with the **Conversion** plugin in QGIS. The vector layer is rasterized according to the

column that now contains the consequence levels (so 4 for very high, 3 for high, 2 for medium and 1 for low).

3. With the Raster calculator, all the raster maps of the assets at risk are multiplied in order to obtain the total consequence built from the intersection of the assets at risk.

Following this procedure gives more insight to just analysing which assets will be flooded in 2100 under a sea level rise of 0,6m. Instead, it provides concise information on the characteristics of the attributes that are flooded, since it is not the same to flood an empty parking area than a naturally protected wetland.

This consequence map gives a detailed information on the valuable assets in order to take them into consideration when designing adaptation measures. Figure 55 shows the results that were obtained.

The map shows that the area of the beach and the northern part, where the wetlands are located, are the most exposed regions. Although it is difficult to perceive from the image, it can also be seen that several urbanised roads along the beach would also have very high consequences, due to its proximity to buildings and businesses.

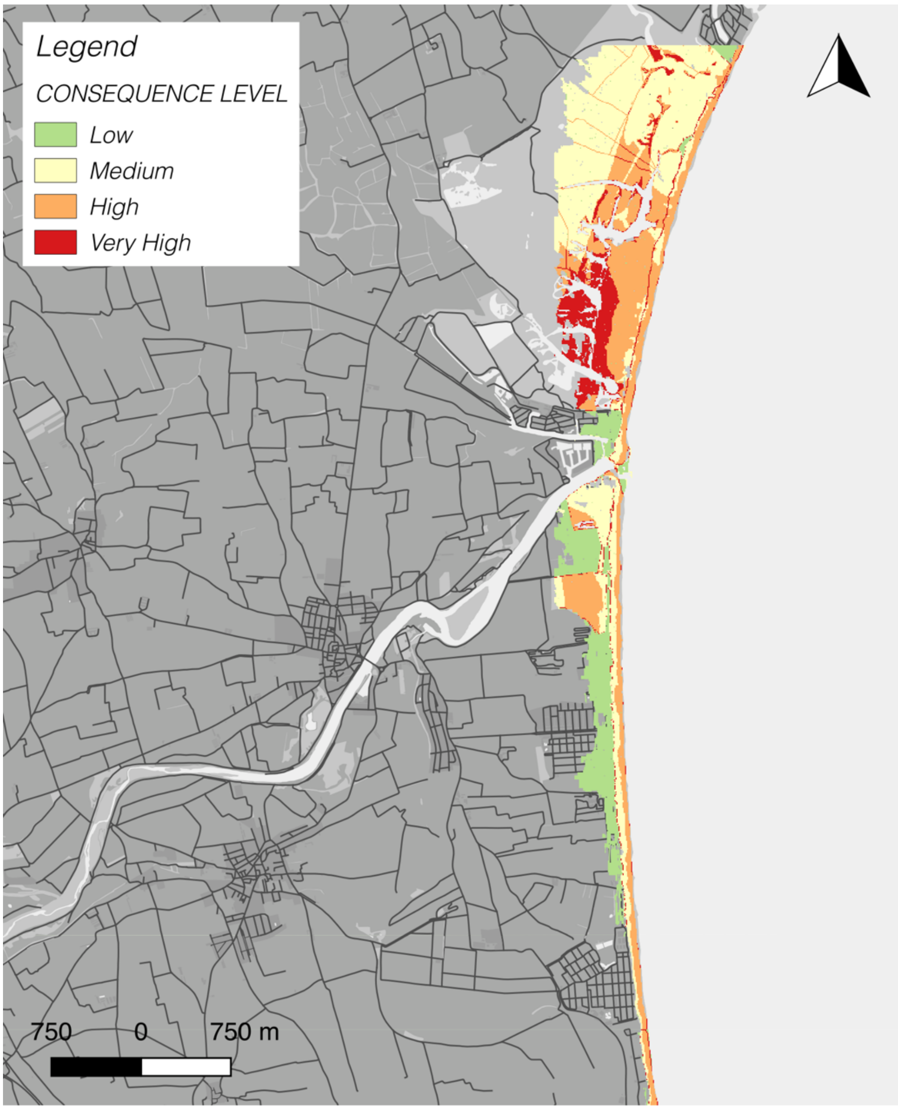


Figure 55 Consequence level map resulting from the intersection of the assets at risk.

9.6.4 Determining the Probability of Occurrence

The probability of occurrence for flooding scenarios can be defined as 1/return period. Since the analysis is considered for a wave return period of 100 years, the probability of occurrence is expected to be 1/100= 1%. However, the use of this value as a definition of the probability of occurrence is normally taken when several flooding scenarios are considered, because the risk analysis results can then be compared between scenarios that are less and more likely to occur.

In this particular case, the risk analysis only focuses on the most severe climate scenario, with a SLR or 0,6 m and a wave return period of 100. This means that the interest lies not that much on how different will each scenario be in terms or risk, but in establishing which are the assets that have to be considered as a priority during the adaptation measures that will have to be set up.

Therefore, to simplify this part, the probability of occurrence will be determined by taking into consideration for the 6 Scenarios, which are the areas that, during flooding events, are flooded. The evolution of the area for the six scenarios is given in Figure 56, and clearly shows that, when flooding occurs, the area that is more likely to flood is the beach coastline and the northern part of the wetland. This means that, flooding of the beach and wetlands is the event that is expected to occur in most circumstances, and more frequently. Therefore, it will be given a higher value of probability.

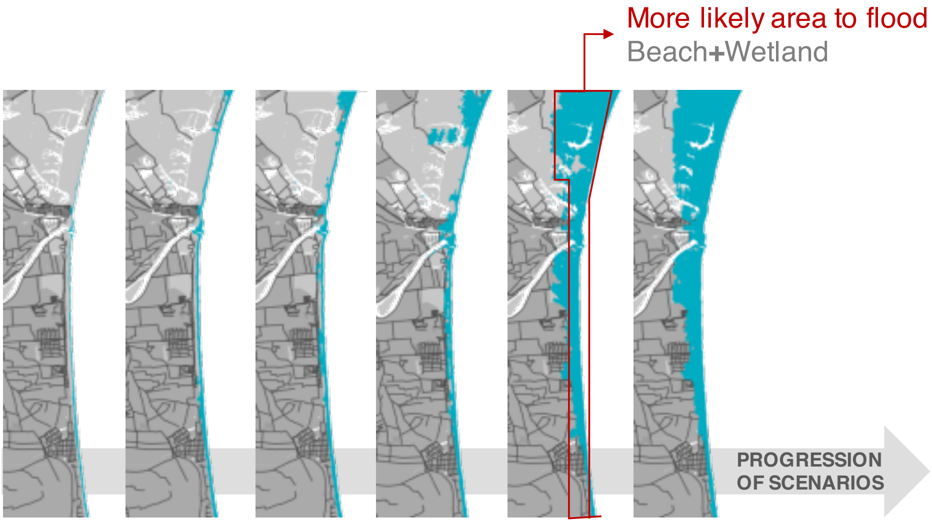


Figure 56 Progression of flooded area that highlights that the beach and the wetlands are the most likely flooded areas due to coastal flooding. Source: Own elaboration

The probability of occurrence will be lower for areas getting further away of the beach and wetlands. In addition, as it can be seen, southern regions close to the coast would also have a smaller probability of occurrence of being flooded (as it has been seen that its morphology acts as a natural barrier.

As a result of that, a hot spot map is obtained, clearly indicating the areas where mitigation and adaptation should concentrate:

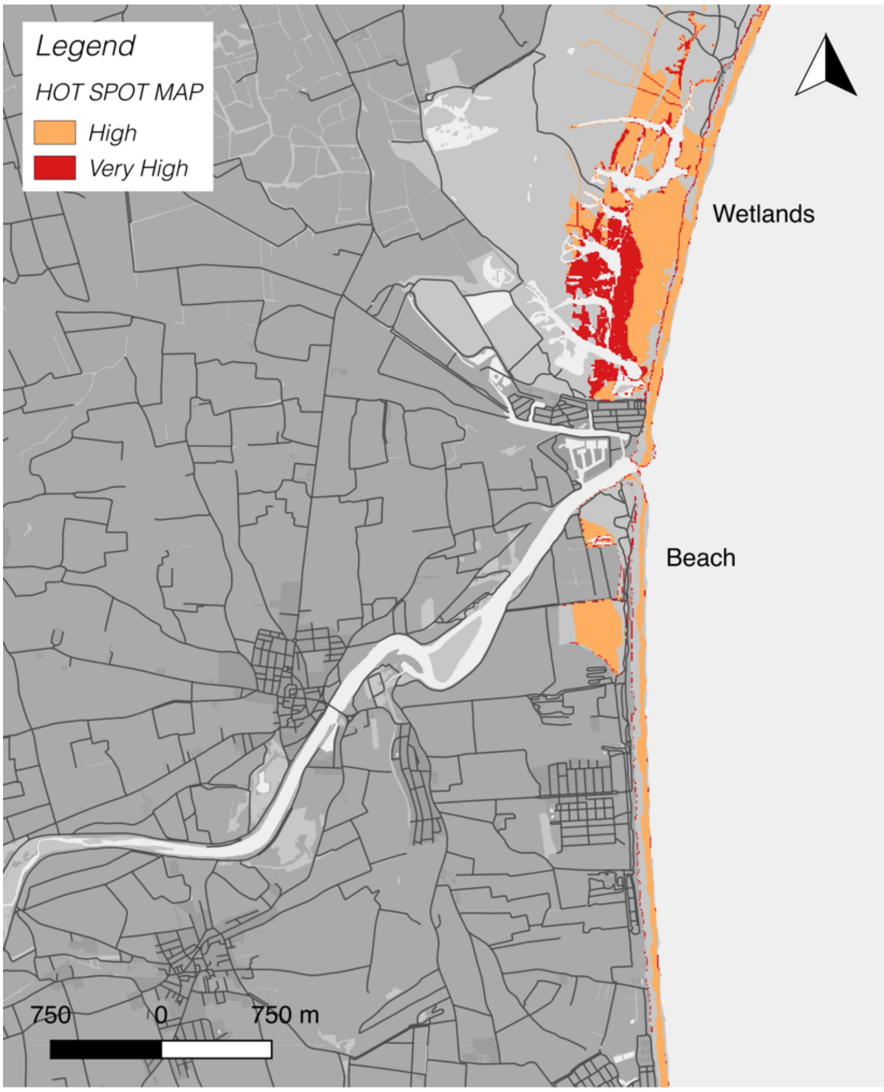


Figure 57 Hot spot areas with a higher Risk of Flooding. Source: Own elaboration

10. Potential actions for Improvement

10.1 Introduction

In coastal management, the generic options for adaptation can be classified in three groups (Zhu et al., 2010).

1. **Protection**: To defend vulnerable areas, especially population centres, economic activities and natural resources.
2. **Accommodation**: To continue to occupy vulnerable areas, but accept the greater degree of flooding by changing land use, construction methods and/or improving preparedness.
3. (**Planned**) **retreat**: To abandon structures in currently developed areas, resettle inhabitants and plan that new development has to set back from the shore.

This last chapter will discuss the potential measures, and which could fit best to the study area.

10.2 Reduction of Risk

The first step is to determine the threshold to be established for preventing coastal flooding. As it has been seen in the las part of the previous chapter, the main objective is to reduce the exposure of the beach and the wetland from flooding, mostly in the northern region. For this, the first objective is to:

- Increase the level of protection in the beach to prevent from major sediment loss in the northern region.
- Reduce the flooding depth reached in the wetlands in order to avoid degradation of a protected natural environment.

The available time to plan and execute the measures has been decided to be before 2040. This is crucial to be taken into consideration because, as it was seen in *7.7 Tipping Points and recommendations*, that would be the year that a sea level rise of 0,25m could be reached under a worst climate change scenario. A sea level rise of concern of 0,25m was set during the logistic regression model, where it was seen that values of sea level rise larger than 0,25m caused flooding water depth of Category 3 (values of water depth between 0,6m and 1m) along the whole beach.

10.3Discussion of measures

The planning and decision of measures depend on the type of coastal system. The characteristics of the beach were described in the study area phase, which showed that is formed by low lying areas and sandy dunes. There are local urbanised areas close-by as well as businesses and several camping.

RISES-AM (***Responses to coastal climate change: Innovative Strategies for high End Scenarios - Adaptation and Mitigation***) is an EU-funded project developed by partners from many countries, including the *Laboratori Enginyeria Maritima,* (LIM/UPC), that described, for the similar type of beach that this present study, which would be the main impacts due to sea level rise (Figure 58).

Type	**Description**	**Main impacts of sea level rise**	**Example**
1a. open, urbanized coast with beach and/or sand dunes 	Urbanised areas, low lying and attractive for tourism. May be protected by sand dunes and sand nourishment to maintain coastline	Erosion of beaches and dunes, damage to tourism. Increased risk of inundation.	Holland coast (the Netherlands), Catalan coast (Spain)

Figure 58 Coastal Zone Archetype Source: RISES-AM

To work in the reduction of the erosion in beaches and dunes, alleviate the damage to tourists and reduce the increase of inundations, several types of measures will be discussed:

The **Protection Approach** aims to protect the shore by eliminating the exposure to flooding, wave attack and erosion (Zhu et al., 2010).
There are two types of measures:

1) Hard Defences: It is the traditional approach and is characterised by building structures to create a barrier between land and sea. The measures can include seawalls, revetments, or sea dikes.
   The main problem of this type of measures is that hard structures fix the position of the coastline, losing the natural dynamic landforms characteristic of these environments. Additional problems exist in the fact that hard structures can impede the recreational use of beaches and can be costly to construct and maintain (USACE, 2002). This is especially critical for the study area since a great part of the land use is for recreational purposes, having more than 9 camping spread along the coast. Therefore, this type of measures will be excluded from the set of possible options, at least for the first years of adaptation measures.
2) Soft Structures: Soft engineering technologies focus on adapting to existing natural processes. In many cases, soft measures can help avoid the negative consequences of hard defences. Therefore, the measures classified as soft structures will be the ones considered and discussed for future planning in the

region. Some examples include beach nourishment, dune building or beach drainage.

10.3.1 Beach nourishment

Beach nourishment is an adaptation technology primarily used in response to shoreline erosion, although flood reduction benefits can also occur. It mainly contributes to wave energy dissipation, done by introducing large quantities of beach material to the coastal sediment budget from an external sediment source, also referred to as a borrow site.

Shoreline erosion will continue to occur, but the widened and deepened beach will provide a buffer to protect coastal infrastructure and other assets from the effects of coastal erosion and storm damage.

The implementation of this measure would be done in the northern region, that is more exposed to flooding and that has less elevation than the southern part, which makes it more vulnerable to flood events. Also, the northern region is also where the wetland is located, so focusing on beach nourishment would prevent water to reach the limits of the wetlands.

Beach nourishment is a flexible coastal management solution. This measure is especially suitable for first climate scenarios, where the sea level is still smaller to 0,25m and most of coastal flooding is caused by wave conditions. As it was seen in this project, for small values of sea level rise, the flood category is mainly affected by wave conditions which means that using beach nourishments for the during the next years could alleviate the impacts in the wave dynamics and storms that will occur in the near future (Zhu et al., 2010).
Nourishment is not a permanent solution to shoreline erosion, so It will require periodic renourishments, or 'top-ups' (or the combination with other methods).

One negative aspect of beach nourishment is that placement of the fill material on the beach can cause a disruption in the ocean and beach habitats. Considering that in the northern region, there is a protected habitat of flora and fauna due to the Wetland, this aspect would require from careful planning and implementation to not interfere with the habitat.

10.3.2 Artificial Dunes and dune rehabilitation

Another measure analysed is the use of artificial dunes. Dunes naturally occur along most undeveloped, sandy coastlines. They can provide both a physical and tangible defence.
Artificial dune construction involves the placement of sediment from dredged sources of the beach. This is followed by reshaping these deposits into dunes using bulldozers or other means (Zhu et al., 2010).

When planning the application, it should be considered that dunes also represent a barrier to beach access, so direct access to the beach could be affected. That would impact the camping accessibility and the local businesses. On the other side, sand dunes also provide a valuable coastal habitat for many highly-specialised plants and animals, so it could be very beneficial to enrich the Wetland Park and its surroundings.

10.3.3 Set-back lines

Apart from protection measures, it is also very important to consider that, due to the tourist demand, many of the businesses and camping are next to the beach, increasing the vulnerability to climate change.
Camping areas are specially exposed to flooding because they concentrate a dense group of people, from local inhabitants to tourists. In addition, campings location attracted businesses to move close-by which put them also at risk.

Future coastal management plan should consider to introduce set-back lines for new constructions in areas close to the camping and the beach, as well as enhancing the protection of camping.

10.4 Planning and Application

RISES-AM suggests a way at which the different adaptation measures could be the implemented during time. Figure 59 shows how would the adaptation pathway look for an area similar to the study area. Indeed, this can be used to assess the changes of the measures in time for the different scenarios. As it can be seen, for values of small sea level rise, the initial stage only contemplates drainage systems and pumps, because there aren't still many flooding events (or not with the level of intensity).

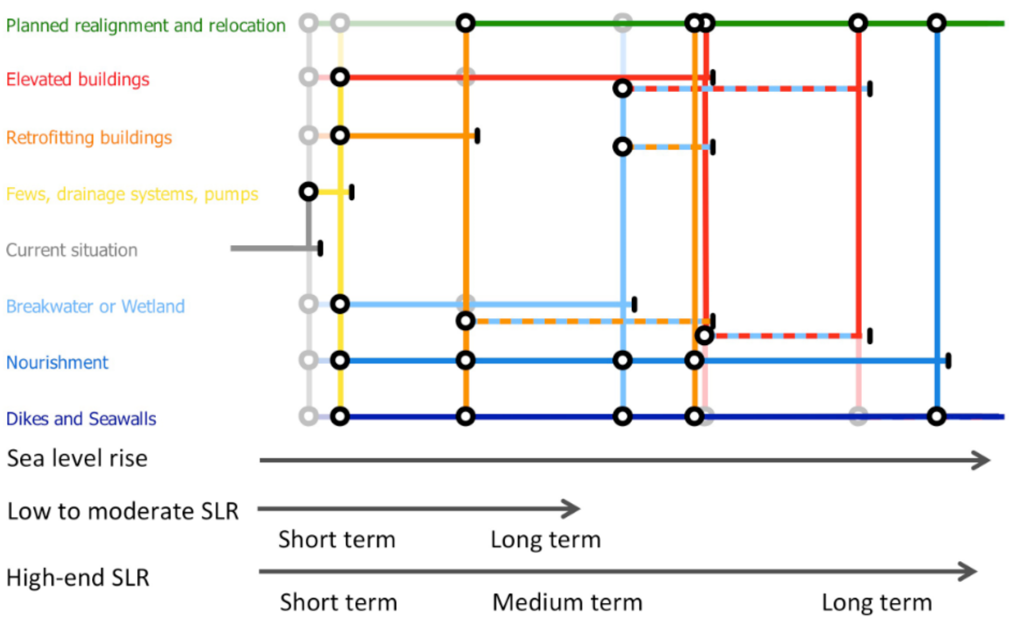


Figure 59 Adaptation Pathways map for a typical Open urbanised coast with beach and sand dunes. Source: RISES-AM

Future planning measures could use the abovementioned methods as well as this Adaptation Pathway map to determine the design of the activities to be performed in the region.

11. Conclusions

This research has contained three main chapters, all of which have tried to cover the central aspects involved in assessing coastal flooding: (a) the development of a mathematical model to predict future flooded scenarios, (b) the analysis of the evolution and trends of flooding in the study area and (c) an integral management and adaptation plan including a risk analysis and tipping points for vulnerability reduction. The main results extracted from this work are detailed in the following chapter.

11.1 Performance of the logistic regression model

Most of the challenges of this project were at the initial phase, during the creation and fitting of the logistic regression model. The data constraints and the limited amount of measurements per points (6 measurements) did not allow to initially split the data in two halves, one for training and the other for testing. Usually, this is a procedure to validate regressions models. This drawback hindered the validation of the model through conventional procedures. In detail, it can be concluded that:

- 6 measurements per point in the map are enough to generate a logistic regression model for predicting coastal flooding categories. However, the results show that sea level rise has far more weight in the category prediction compared to the Wave Return Period.
- 7,5% of the study points were excluded from the analysis because they did not present data variability (all the measurements had the same category), which disqualified them for the logistic model fitting.
- Given the small amount of available data, extreme probabilities were obtained for the category that was predicted, with values of, for example, 0,999959545 for a given category and values of expected probability of the other categories of the order of 10^{-5}.
- For the smallest prediction scenarios (where the Sea Level Rise is 0), the complete or quasi complete separation accounts for more than 10% of the total points. This means that the predictor (or a linear combination of some subset of the predictors) is associated with only one outcome value, and therefore can not associate a certain probability because the model itself has not been applied to predict the category.
- Concerning the software used, NNET package in R, does not include a warning information on the phenomenon of complete or quasi-complete separation (opposed to other packages that also perform logistic regression analysis like *GLMNET*).

Hence, it can be concluded that working with a point-by-point based regression model requires more input information, not only consisting of 6 measurements. Mainly, because of the amount of information that is lost in the model creation.

Since the conventional validation methods were not able to be carried out (mostly due to the lack of data), a new model consisting of polygons that grouped the points was carried out, in order to serve as a comparison and validation with the point-by-point model.

11.1.1 Validation of the Model – Node-based vs Polygon Based

- Building the logistic regression with polygons instead of points results in a higher number of measurements per model (since each polygon contains the measurements of several points). As a consequence, the 7,5% losses accounted for data variability for the point-by-point model are reduced to 0,5%.
- A polygon-based model consisting of several points does not reduce the accuracy of the predictions of category compared to the point-by-point model. Almost 100% of the polygons analyzed predict the same category as the most represented category by the points included within the polygon.
- A polygon-based model is particularly beneficial for studying probabilities. It has more homogeneous probability results that allow to give an estimate on the potential of higher and lowed categories to appear in a future scenario. This is considered to be highly valuable for vulnerability reduction measures and to highlight hot spot areas.

11.2 Discussion on Hydrodynamic models. Sensitivity analysis Thresholds and Tipping Points

After applying the logistic regression model for different future scenarios of sea level rise and wave return period, it can be concluded that:

- For Predictions 1 to 3 (with no sea level rise), the vast majority of points are comprised between category 1 and 2, whereas for higher predictions the major category is 3.
- Comparing the spatial distribution of categories alongside the study area, it can be concluded that the Norther Region (delimitated by Zone 1, 2 and 3) is more prone to flooding (and more likely to reach higher categories). Under large predictions of sea level rise and return period, close to 20% of the points in the Northern region are in Category 4 whereas the prediction scenario for the Central and Southern region contains less than 5% of the points in Category 4.
- The category predicted is strongly influenced by the SLR and not much by the return period. When a sea level rise of 0,6 m is introduced, there is a sharper increase of the category compared to the predictions when there is only an increase in the return period.
- Comparing the probability to get a certain Category for the predictions that have a Sea Level Rise estimation of 0 (Prediction 1,2 and 3), a category 1 probability is larger than 75%, with a remaining 10% of probability to obtain a Category 2. For the cases where there is a sea level rise of 0,6m, there is a 100% of probability to get categories higher than 1. As previously mentioned, the category with higher probability to occur is Category 3 (with 70% of probability).

11.2.1 Concerning the Sensitivity analysis:

- For changes in Sea Level Rise below 0,25 m, there are no significant differences in the number of points in categories 1, 2, 3 and 4. There is a tendency, for low wave return period to have 1 as the main category whereas for a large wave return period of 75 years the main category is 2.
- For changes in Return Period combined with small values of Sea Level Rise, the return period is accurately predicting a similar number of points in category 1, 2, 3 and 4 for wave return period below 25. If the Sea Level Rise takes larger values, this regular behavior is not followed anymore. This helps to conclude that the Sea Level Rise is the variable that has a higher impact on the outcome and that, for high values, the return period does not influence that much in the expected category.
- For high values of sea level rise, the number of points in category 2 remains constant whereas the number of points category 1 continue to reduce considerably while category 3 points increase. This reinforces the hypothesis that for small increases in sea level rise, the effect on the overall category is very strong
- Sea level rise threshold: As shown in Figure 29 and
- Figure 30, most of the changes in the number of points per category is experienced for values larger than 0,25 m of sea level rise. For values of sea level rise smaller than 0,25m, the wave return period is the important variable for coastal flood prediction.
- For small values of Sea Level Rise (0,05 - 0,15m), the category increases gradually shifting from Category 1 to Category 2. For high sea level increases over a certain threshold close to 0,25m – 0,30m, the Category levels move directly to Category 3 without transitioning through category 2 scenarios.
- According to predictions, a sea level rise of 0,25m will be reached in a time range of 2040 – 2070 (under aggressive climate scenario), leaving only a window between 5 and 10 years more to reach it (under less aggressive climate scenarios). Meeting the threshold of 0,25m of SLR will be translated to a Category 3 damage over the study area extension (water depth from 0,6 – 1m).

11.3 Assets at risk and management application

Transportation facilities, vegetation and land cover, water resources and building infrastructure will be highly affected by coastal flooding:

- Over 9 campings around the study area will be exposed to the worst-case Scenario, threatening as well the adjacent businesses and local restaurants.
- More than 20 km of secondary non-paved roads and rural paths are expected to flood in the scenarios where sea level rise reaches 0,6m.
- The Natural Wetland Park of l'Empordà, a protected environment currently declared national natural interest due to its integral zoological and botanical reserves, will suffer from severe flooding, even before a 0,6m sea level rise is reached.
- Several kilometres of water streams of closed water bodies will also be flooded by water, altering the water quality with saline intrusions, sediments and marine pollutants. Over 25 Km of streams, rivers and water masses will be reached by seawater in the scenarios with sea level rise of 0,6m.

- The risk analysis shows that the Sant Pere Pescador Beach and the Wetland Park of l'Empordà will be the most vulnerable assets in 2100, suffering from prolonged flooding way early in advance before the worst-case scenario is reached.
- The 3D representation of the Digital Elevation Model shows that the southern region has higher elevation than the northern region, as well as more roughness in the beach given by the sandy dunes which acts as a natural barrier for coastal flooding (since it is the region less flooded of the map. Opposed to it, the surface in the northern region is more homogeneous and in lower elevation, facilitating seawater to impact the beach.

11.4 Potential recommendations

- During the first years, alleviate the impacts of flooding by draining measures and pumps, since the flooding is still very localized and depends and takes place
- In the coming years, plan to develop protection measures mostly to the northern region and along the beach, by implementing beach nourishment and artificial dunes. An increase in the amount of sand in northern regions would reduce wave energy impact, mitigating the amount of seawater flooding to the wetlands.
- Integrate in the current coastal management plan, the assets at risk and the stakeholders that would be affected by coastal flooding, in order to agree on a common adaptation strategy to reduce not only the negative impacts on the coast and its natural environment, but also the consequences for businesses and services offered in the area.

References

Booij, N., Ris, R.C. y Holthuijsen, L.H. (1999): A third-generation wave model for coastal regions, part I: model description and validation. J. Geophys. Res., 104 (C4): 7649-7666.

Bucx, T., Van der Brugge, R., Haasnoot, M., (2017) Effective interventions and adaptation strategies for coastal zones under climate change. Policy Brief. RISES-AM.

Centre Internacional d'Investigació dels Recursos Costaners, CIIRC. (2010): Estat de la zona costanera a Catalunya, Vol I, II, III y IV. Barcelona. Generalitat de Catalunya.

Church, J. a., Clark, P. U., Cazenave, a., Gregory, J. M., Jevrejeva, S., Levermann, a., ... Unnikrishnan, a. S. (2013). Sea level change. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 1137–1216. https://doi.org/10.1017/CB09781107415315.026

Doornkamp, J. C. (1998). Coastal flooding , global warming and environmental management. *Journal of Environmental Management*, *52*(March 1997), 327–333. https://doi.org/10.1006/jema.1998.0188

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2003). *Applied Logistic Regression. 3rd edition. Technometrics*. https://doi.org/10.2307/1270433

Institute for Digital Research and Education (IDRE). UCLA: Statistical Consulting Group, 2017. From http://stats.idre.ucla.edu/other/multpkg/faq/general/-faqwhat-is-complete-or-quasi-complete-separation-in-logisticprobit-regression-and-how-do-we-deal-with-them/ (accessed May 15, 2017).

Losada, I., Izaguirre, C. & Diaz, P. 2014. Cambio climático en la costa española. Oficina Española de Cambio Climático, Ministerio de Agricultura, Alimentación y Medio Ambiente. Madrid.

Muis, S., Güneralp, B., Jongman, B., Aerts, J. C. J. H., & Ward, P. J. (2015). Flood risk and adaptation strategies under climate change and urban expansion: A probabilistic analysis using global data. *Science of the Total Environment*, *538*, 445–457. https://doi.org/10.1016/j.scitotenv.2015.08.068

National Oceanic and Atmospheric Administration, 2015. United States Office of Coast Survey, Coast Survey Development Laboratory http://www.stormsurge.noaa.gov-/overview_causes.html (accessed May 14, 2017).

Neyman, J., Pearson, E.S. (1933), On the Problem of the most efficient tests of statistical hypotheses, Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences, 231 (694-706): 289-337

NSW Treasury, (2012). Risk Management Toolkit for NSW Public Sector Agencies. Volume 1: Guidance for agencies

Pinyol, J., García-león, M., González, M., Lin-Ye, J., Oller, P., Gracia, V., & Sánchez-Arcilla, A. (2015). Metodología para la delimitación de la zona de inundación del oleaje en el Mapa para la Prevención de Riesgos Geológicos de Cataluña Coastal flooding assessment on the Map for the ..., (May 2016), 2007–2010.

Purvis, M. J., Bates, P. D., & Hayes, C. M. (2008). A probabilistic methodology to estimate future coastal flood risk due to sea level rise. *Coastal Engineering*, *55*(12), 1062–1073. https://doi.org/10.1016/j.coastaleng.2008.04.008

Roelvink D., Reniers A., van Dongeren A., van Thiel de Vries J., McCall R. y Lescinski J. (2009): Modeling storm impacts on beaches, dunes and barrier islands, Coast. Eng., 56, Issues 11-12: 1133- 1153.

Sánchez-Arcilla, A., García-León, M., and Gracia, V.: Hydro-morphodynamic modelling in Mediterranean storms – errors and uncertainties under sharp gradients, Nat. Hazards Earth Syst. Sci., 14, 2993-3004, doi:10.5194/nhess-14-2993-2014, 2014.

Sánchez-Arcilla, A., García-León, M., Gracia, V., Devoy, R., Stanica, A., & Gault, J. (2016). Managing coastal environments under climate change: Pathways to adaptation. *Science of the Total Environment*, *572*, 1336–1352. https://doi.org/10.1016/j.scitotenv.2016.01.124

Sánchez-Arcilla, A., Gracia, V., Sierra, J. P., García-León, M., & Mösso, C. (2016). *Tercer informe sobre el canvi climàtic a Catalunya: 8. Sistemes Costaners i Dinàmica Litoral*. Retrieved from http://static-m.meteo.cat/wordpressweb/wp-content/uploads/2017/01/30132416/3-TICCC-Cap_4-Evolucio-recent-de-la-temperatura-la-precipitacio-i-altres-variables-climatiques-a-Catalunya.pdf

Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.), 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Van Vuuren et al, 2011. The Representative Concentration Pathways: An Overview. Climatic Change, 109 (1-2), 5-31

Wilks, S.S. (1938), The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses. Annals of Mathematical Statistics, 9, 60-62

Wong, P. P., Saito, Y., Sallenger, A., Losada, I. J., Gattuso, J.-P., McInnes, K. J., ... Khattabi, A. (2013). Coastal Systems and Low-Lying Areas. *Climate Change 2014 Impacts, Adaptation, and Vulnerability*, (October 2013), 361–410. https://doi.org/10.1017/CBO9781107415379.010

Yang, Y. C. E., Ray, P. A., Brown, C. M., Khalil, A. F., & Yu, W. H. (2014). Estimation of flood damage functions for river basin planning: a case study in Bangladesh. *Natural Hazards*, *75*(3), 2773–2791. https://doi.org/10.1007/s11069-014-1459-y

Zhu, X., Linham, M. M., & Nicholls, R. J. (2010). Technologies for Climate Change Adaptation - Coastal Erosion and Flooding. Roskilde: Danmarks Tekniske Universitet, Risø Nationallaboratoriet for Bæredygtig Energi. (TNA Guidebook Series).

Zou, Q. P., Chen, Y., Cluckie, I., Hewston, R., Pan, S., Peng, Z., & Reeve, D. (2013). Ensemble prediction of coastal flood risk arising from overtopping by linking meteorological, ocean, coastal and surf zone models. *Quarterly Journal of the Royal Meteorological Society*, *139*(671), 298–313. https://doi.org/10.1002/qj.2078